The American Economic Review

ARTICLES

CUK-H01165-97-P004419

391
.001

97

KATHRYN M. DOMINGUEZ, RAY C. FAIR, AND MATTHEW D. SHAPIRO
Forecasting the Depression: Harvard versus Yale

PAUL R. GREGORY AND IRWIN L. COLLIER, JR.
Unemployment in the Soviet Union: Evidence from the Soviet Interview Project

LEIF DANZIGER
Costs of Price Adjustment and the Welfare Economics of Inflation and Disinflation

GUILLERMO A. CALVO
Servicing the Public Debt: The Role of Expectations

LAURENCE J. KOTLIKOFF, TORSTEN PERSSON, AND LARS E. O. SVENSSON
Social Contracts as Assets: A Possible Solution to the Time-Consistency Problem

ZOLTAN J. ACS AND DAVID B. AUDRETSCH
Innovation in Large and Small Firms: An Empirical Analysis

CHARLES BLACKORBY AND DAVID DONALDSON
Cash versus Kind, Self-Selection, and Efficient Transfers

WALTER HETTICH AND STANLEY L. WINER
Economic and Political Foundations of Tax Structure

JENNIFER F. REINGANUM
Plea Bargaining and Prosecutorial Discretion

ARYE L. HILLMAN AND HEINRICH W. URSPRUNG
Domestic Politics, Foreign Interests, and International Trade Policy

LEONARD K. CHENG
Assisting Domestic Industries under International Oligopoly: The Relevance of the Nature of Competition to Optimal Policies

ROY J. RUFFIN
The Missing Link: The Ricardian Approach to the Factor Endowments Theory of Trade

RICHARD E. BALDWIN
Hysteresis in Import Prices: The Beachhead Effect

SHORTER PAPERS: W. H. Kaempfer and A. D. Lowenberg; S. Ishizawa, A. Alesina; A. E. Roth, J. K. Murnighan, and F. Schoumaker; J. Neelin, H. Sonnenschein, and M. Spiegel; K. Binmore, A. Shaked, and J. Sutton.

SEPTEMBER 1988

THE AMERICAN ECONOMIC ASSOCIATION

• Printed at Banta Company, Menasha, Wisconsin.

• No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

Correspondence relating to advertising, business matters, permissions to quote, subscriptions, and changes of address, should be sent to the American Economic Association, 1313 21st Avenue South, Suite 809, Nashville, TN 37212–2786. Please remit membership payment with the application included elsewhere in this journal. Change of address notice must be received at least six (6) weeks prior to the publication month. A membership or subscription paid twice is automatically extended for an additional year unless otherwise requested.

THE AMERICAN ECONOMIC REVIEW (ISSN 0002–8282), September 1988, Vol. 78, No. 4, is published five times a year (March, May, June, September, December) by the American Economic Association, 1313 21st Avenue South, Suite 809, Nashville, TN 37212–2786. Annual subscription fees: Institutional subscriber-$125.00, Individual subscriber-$72.00, Regular members-$38.50, $46.20, or $53.90 depending on income. A subscription also includes the *Journal of Economic Literature* and the *Journal of Economic Perspectives*. In countries other than the U.S.A., add $16.00 for extra postage. Second-class postage paid at Nashville, TN and at additional mailing offices. POSTMASTER: Send address changes to the *American Economic Review*, 1313 21st Avenue South, Suite 809, Nashville, TN 37212–2786.

Founded in 1885

Officers

President
ROBERT EISNER
Northwestern University

President-elect
JOSEPH A. PECHMAN
The Brookings Institution

Vice-Presidents
MARTIN S. FELDSTEIN
National Bureau of Economic Research and Harvard University
F. M. SCHERER
Swarthmore College

Secretary-Treasurer
C. ELTON HINSHAW
Vanderbilt University

Editor of The American Economic Review
ORLEY C. ASHENFELTER
Princeton University

Editor of The Journal of Economic Literature
JOHN PENCAVEL
Stanford University

Editor of The Journal of Economic Perspectives
JOSEPH E. STIGLITZ
Stanford University

Executive Committee

Elected Members of the Executive Committee
SHERWIN ROSEN
University of Chicago
THOMAS J. SARGENT
University of Minnesota
ROBERT J. BARRO
Harvard University
JUDITH A. THORNTON
University of Washington
GEORGE A. AKERLOF
University of California-Berkeley
ISABEL V. SAWHILL
Urban Institute

EX OFFICIO Members
ALICE M. RIVLIN
The Brookings Institution
GARY S. BECKER
University of Chicago

391
1001

THOMAS CROMBIE SCHELLING

DISTINGUISHED FELLOW

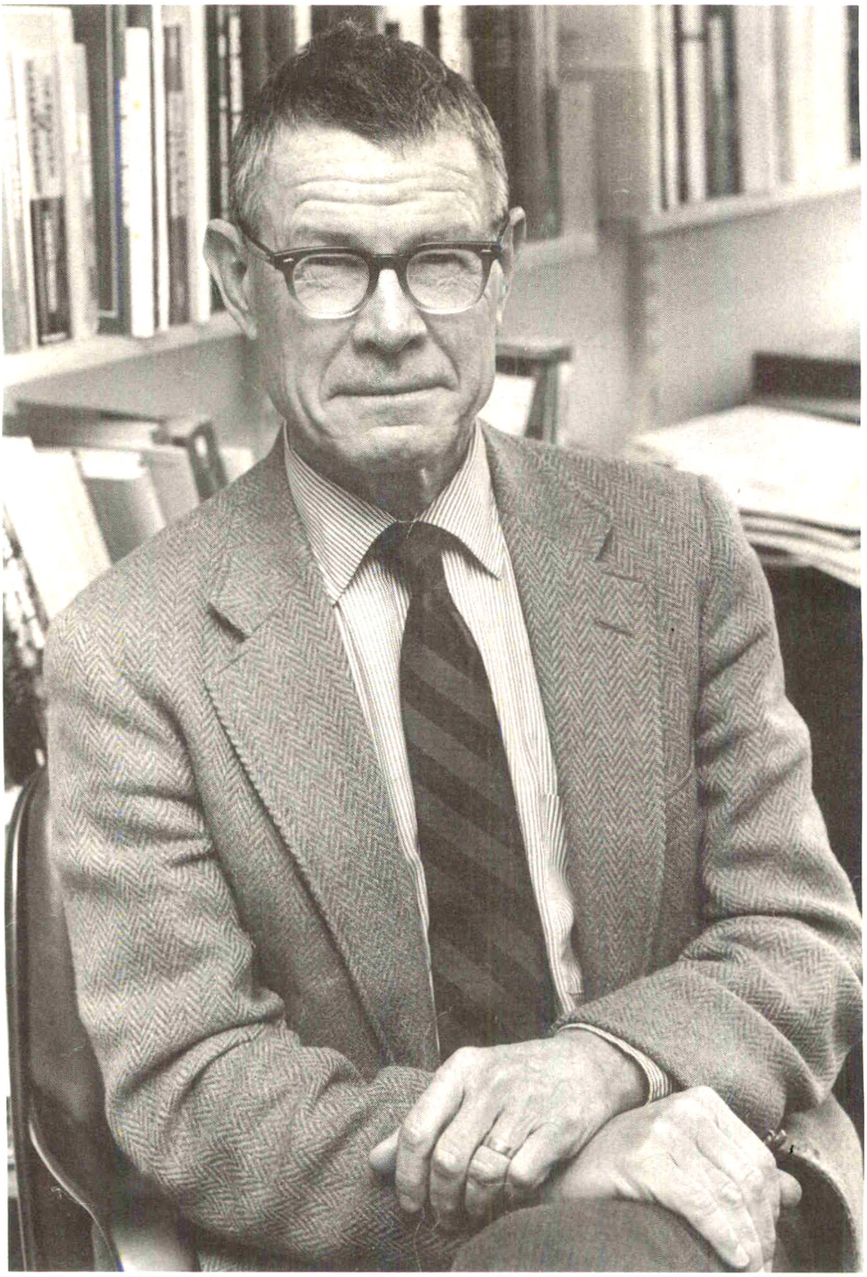
1988

Thomas Crombie Schelling pioneered in the field of "nonmathematical game theory,"—the analysis of social and economic behavior when each participant's behavior depends on his expectations about how each of the others will respond to his actions, taking account of the fact that each of the others is similarly influenced by expectations about how he will respond to their behavior. Schelling erected a conceptual framework for describing such morasses of reciprocal expectations, and by so doing unified a wide variety of disjoint theories, including the theories of oligopoly, military and diplomatic strategy, negotiation, teamwork, and communication.

Schelling's theory of social relationships and his varied applications of it result from a uniquely fruitful symbiosis of theory and practice. He has an extraordinary gift for perceiving the essence of a social or economic situation involving several participants of diverse or coincident interests, and portraying it in vivid, concrete terms. He has presented his theoretical insights and their practical applications in a long series of influential papers and monographs.

As a teacher, he has excited generations of students in his classic course on "Conflict and Strategy," enriched by practical illustrations, and has made his theories widely accessible by many lively and stimulating articles addressed to the general reader.

P 4419



TcSchelling

THE AMERICAN ECONOMIC REVIEW

Editor

ORLEY ASHENFELTER

Co-Editors
ROBERT H. HAVEMAN
JOHN B. TAYLOR
HAL R. VARIAN

Production Editor
CLAIRE H. COMISKEY

Board of Editors

GEORGE A. AKERLOF
JAMES E. ANDERSON
JO ANNA GRAY
GEORGE E. JOHNSON
KENNETH L. JUDD
JOHN F. KENNAN
MAURICE OBSTFELD
ROBERT H. PORTER
JOHN G. RILEY
RICHARD ROLL
ALVIN E. ROTH
DAVID SAPPINGTON
KENNETH J. SINGLETON
ROBERT S. SMITH
BARBARA J. SPENCER
LESLIE YOUNG

- Submit manuscripts (4 copies), 50 pages maximum, single-sided, double-spaced, to:
  Orley Ashenfelter, Editor, *AER*; 209 Nassau Street, Princeton, NJ 08542–4607.

- Submission fee: $50 for members; $100 for nonmembers. Please pay with a check or money order payable in United States Dollars. Canadian and Foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Style guides will be provided upon request.

- Copyright © American Economic Association 1988. All rights reserved.

September 1988

VOLUME 78, NUMBER 4

Articles

Forecasting the Depression: Harvard versus Yale *Kathryn M. Dominguez, Ray C. Fair, and Matthew D. Shapiro* 595

Unemployment in the Soviet Union: Evidence from the Soviet Interview Project *Paul R. Gregory and Irwin L. Collier, Jr.* 613

Costs of Price Adjustment and the Welfare Economics of Inflation and Disinflation *Leif Danziger* 633

Servicing the Public Debt: The Role of Expectations *Guillermo A. Calvo* 647

Social Contracts as Assets: A Possible Solution to the Time-Consistency Problem *Laurence J. Kotlikoff, Torsten Persson, and Lars E. O. Svensson* 662

Innovation in Large and Small Firms: An Empirical Analysis *Zoltan J. Acs and David B. Audretsch* 678

Cash versus Kind, Self-Selection, and Efficient Transfers *Charles Blackorby and David Donaldson* 691

Economic and Political Foundations of Tax Structure *Walter Hettich and Stanley L. Winer* 701

Plea Bargaining and Prosecutorial Discretion *Jennifer F. Reinganum* 713

Domestic Politics, Foreign Interests, and International Trade Policy *Arye L. Hillman and Heinrich W. Ursprung* 729

Assisting Domestic Industries under International Oligopoly: The Relevance of the Nature of Competition to Optimal Policies *Leonard K. Cheng* 746

The Missing Link: The Ricardian Approach to the Factor Endowments Theory of Trade *Roy J. Ruffin* 759

Hysteresis in Import Prices: The Beachhead Effect *Richard E. Baldwin* 773

Shorter Papers

The Theory of International Economic Sanctions: A Public Choice Approach
William H. Kaempfer and Anton D. Lowenberg 786
Increasing Returns, Public Inputs, and International Trade *Suezo Ishizawa* 794
Credibility and Policy Convergence in a Two-Party System with Rational Voters
Alberto Alesina 796
The Deadline Effect in Bargaining: Some Experimental Evidence
Alvin E. Roth, J. Keith Murnighan, and Francoise Schoumaker 806
A Further Test of Noncooperative Bargaining Theory:
Comment *Janet Neelin, Hugo Sonnenschein, and Matthew Spiegel* 824
Reply *Ken Binmore, Avner Shaked, and John Sutton* 837

Preliminary Announcement of the AEA Meeting **841**

Forecasting the Depression: Harvard Versus Yale

By KATHRYN M. DOMINGUEZ, RAY C. FAIR, MATTHEW D. SHAPIRO*

Was the Depression forecastable? After the Crash, how long should it have taken contemporary forecasters to realize how severe the downturn was going to be? Data assembled by the Harvard and Yale forecasters — together with modern historical data — are subjected to statistical analysis. Neither contemporary forecasters nor modern times-series analysts could have forecast the large declines in output following the Crash.

James Tobin relates the following story, which was told to him and other members of the Harvard graduate economics club by Professor W. L. Crum. In the 1920s the Harvard Economic Service (HES) issued monthly reports on the current and expected future state of the economy. HES used three indexes, representing speculation, business, and money, to help predict the future. Crum claimed that in the summer of 1929 the statistical assistant at HES became alarmed when she noticed that the indexes indicated that a sharp downturn in economic activity was imminent. Crum did not see in the current business situation any cause for this adverse forecast. Moreover, he feared that a pessimistic forecast by the influential service could itself have an adverse effect on financial markets and economic activity. Therefore, he suppressed the pessimistic findings of the assistant;[1] the published report did not speak of a potential downturn.

The data, in fact, provide only mild support for this account. The account does, however, raise the more general question of whether the Depression was forecastable. Was there anything in the data prior to the October 1929 stock market Crash that indicated the economy was about to enter a protracted slowdown? How should the news of the Crash have revised forecasts of economic growth? To address these questions, we study two sets of data assembled by contemporary business-conditions forecasters. The first consists of the three HES indexes. The second consists of commodity and stock price indexes compiled and analyzed by Irving Fisher. Fisher, who was at Yale from 1891 until his death in 1947, was a competitor of HES. He also released periodic reports on the state of the economy, and he was a critic of the HES indexes.

We also use modern historical data on industrial production, producer prices, stock prices, time-loan rates, and the money stock in our statistical analysis. The use of these data allows us to address the broader question of the forecastability of the Depression without recourse to the services' idiosyncratic data. We can also ask whether the Depression-era data contain information not embodied in the modern data.

The procedures, data, and pronouncements of the Harvard and Yale forecasting services are discussed in Section I, where the accuracy of Crum's account is also investigated. Both services failed to anticipate the Depression and remained optimistic about economic performance following the Crash. In the following section, we examine the data using time-series analysis. The analysis suggests that the services' optimism was not unwarranted.

---

*Dominguez: John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138. Fair and Shapiro: Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520, and National Bureau of Economic Research. We are grateful for helpful comments from James Tobin, Paul A. Samuelson, and anonymous referees.

[1] And consequently he lost the opportunity to gain a reputation for Delphic wisdom.

I. The Harvard and Yale Forecasting Services

The Harvard and Yale economic forecasting services were probably the two preeminent economic analysis and forecasting services available to businesses and members of the general public in the 1920s. Despite their similarities—their association with prominent academic institutions—they were quite different in purpose and organization. The Harvard service, which had a full-time professional staff, was designed as a business whose object was to provide corporations with economic analysis for short- and long-range planning. By contrast, Fisher provided a service that was designed originally as an educational tool for the public. Later, Fisher developed a full-scale advisory business,[2] but initially his pronouncements were published weekly in a syndicated newspaper column on economic affairs that appeared nationwide.

This section summarizes the activities of the two forecasting services. The Appendix gives a more detailed chronology of their pronouncements.

A. *Harvard Economic Service*

The HES forecasts, which were offered to businesses by annual subscription,[3] were centered on the Service's Index of General Business Conditions, which was created by HES's first editor, Warren Persons. Its three curves were meant to represent "speculation" (the A-curve), "business" (the B-curve), and "money" (the C-curve). The Harvard forecasts were then based on the relations determined to exist among the three curves during any given phase of the business cycle and on the magnitude of the movement from peak to trough of each curve (Persons, 1922).

Although the specific series that made up the three curves changed over the thirteen years that HES published the indexes, the basic index methodology remained the same. Curve A, representing speculation, included a series of New York bank clearings and industrial stock prices; curve B, representing business, included outside bank debits and commodity prices;[4] curve C, representing money, was based upon commercial paper rates.

The construction of the three indicators was as follows. First, the underlying component(s) in the index were adjusted for seasonality. Then, for each underlying component, a "similar" series *not* in the index was isolated, which was thought to represent a secular trend relative to the cyclical component in question. (For example, for the "money" series, the baseline was the average yield of ten prime railroad bonds of distant maturities.) Percentage deviations between each component of the index and the "baseline" series were taken and then normalized by the appropriate standard deviation. The resulting groups of adjusted components were then averaged to form the relevant index curve.

The resulting indexes are illustrated in Figures 1 and 2 for, respectively, the years 1904–13 and 1919–31.[5] The Harvard data

---

[2] In 1930, Fisher began to market the *Financial Analysis Service*, published by the Index Number Institute (a company founded by Fisher), which included analysis and economic forecasts based upon the Fisher indices. This publication was succeeded by a similar service named *Trade and Money Index* in late 1931. This service was in turn succeeded by *Market Indicators*, which ceased publication in 1934.

[3] A subscription to the service cost $100 per year ($805 in 1987 dollars) and consisted of the *Weekly Letter*, which contained the Index of General Business Conditions and an accompanying analysis and forecast, the *Quarterly Review of Economic Statistics*, and special statistical supplements.

[4] Curve B as first constructed by Persons in 1919 was a composite of economic series that reflected business conditions, including both production of materials and goods (most notably, pig-iron production) and transactions in commodities and services. The curve was revised in 1923 to include only bank debits and a ten-commodity price index. The position of the baseline (Bradstreet's index) used in constructing curve B "was determined by shifting the curve upward so that the crossing points with the ten-commodity index would correspond to the dates of crossing given for pig iron production, the index of trade, and outside debits." [Persons, 1923, p. 187]

[5] As can be seen, there appears to be strong predictive power in the earlier period. A close inspection of HES's brochure suggests, however, that the remarkable fit was in fact achieved by an "exhaustive study of business statistics, for the primary purpose of developing a reliable index of general business conditions" [Harvard University Committee on Economic Research, 1923b, p. 8] and that this exhaustive study was performed using data from the years 1904–13.

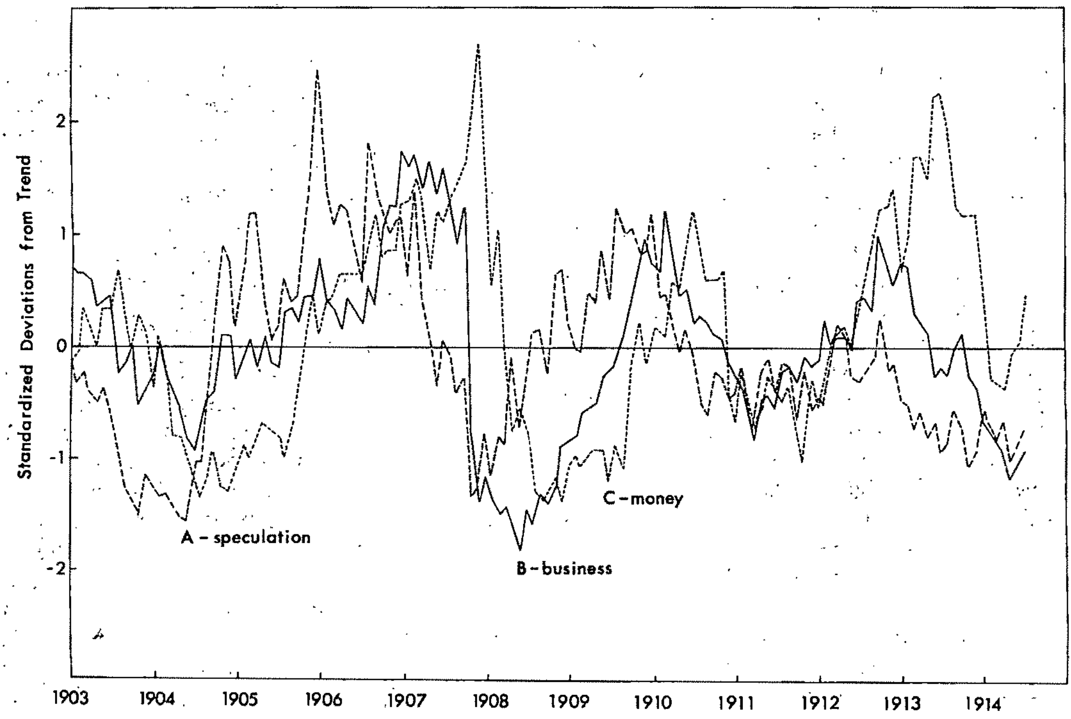


FIGURE 1. HES TEST-PERIOD INDEXES, 1903–14

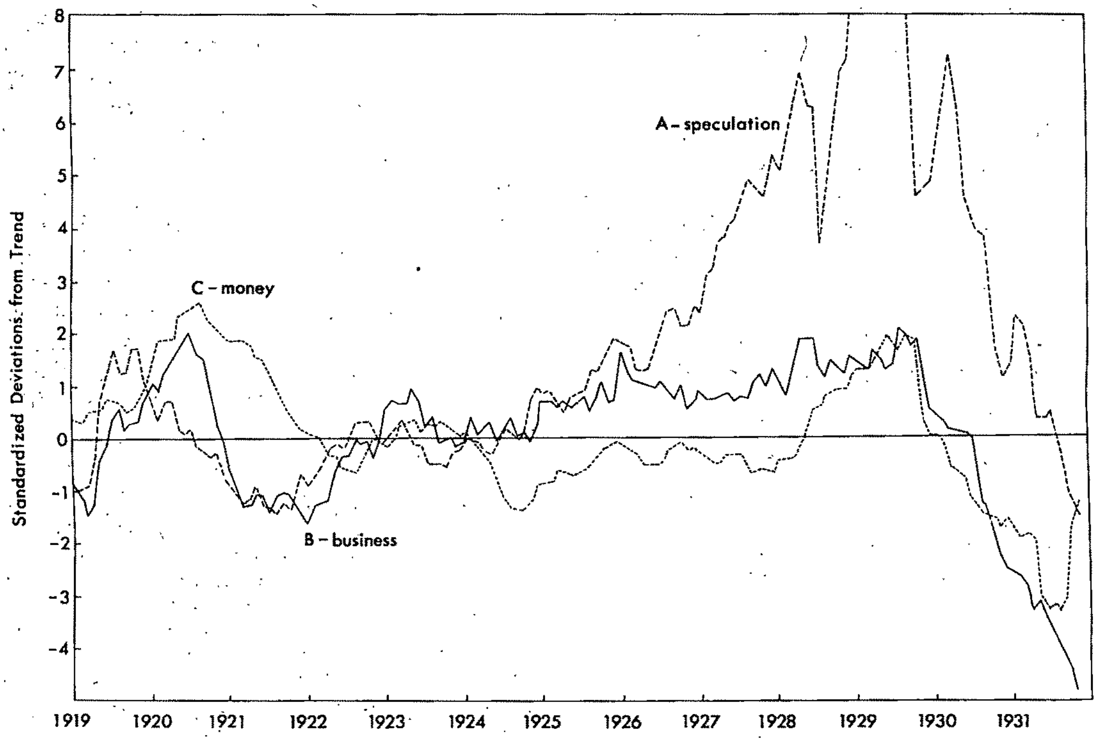


FIGURE 2. HES INDEXES, 1919–31

do not appear at first glance to support Crum's story that the HES indicators forecast a recession in the summer of 1929. All three indexes appear to lag the Crash, rather than anticipate it. Close reading of the HES brochure suggests, however, that it is possible that the statistical assistant foresaw a collapse or at least a downturn at this point. The brochure states that "the index [*C*] forecasts speculation [*A*]" and that "a persistent rise in interest rates was the forerunner of a decline in security prices; after such a decline in security prices, a decline in business [*B*] of several months duration...." [Harvard University Committee on Economic Research (1923b), p. 11).] In the period 1927–29 curve *C* had reflected exactly such a persistent rise in interest rates, and while it is not clear from the service's description exactly when such a rise should ultimately be expected to imply a stock market downturn, it is entirely possible that by the summer of 1929 the assistant saw a market decline as imminent. More importantly, the speculation index (*A*) shows a major decline during the second quarter of 1929. It peaked in March and by June had declined by over 10 percent. This major but transitory decline is almost lost in a graph of curve *A* including the Crash, but it would have looked substantial during the second quarter.[6]

If at least some on the Harvard staff believed that the data foretold business doom, however, this was not apparent in the HES's public pronouncements through the period of collapse. A detailed chronology of these pronouncements is included in the Appendix. The service painted a fairly gloomy picture earlier in 1929, only to moderate its view as the stock market continued to boom. Immediately after the stock market Crash HES pointed to the fact that interest rates had promptly declined, indicating, in their view, the soundness of Federal Reserve policy. Confidence in an imminent recovery continued until late 1931 when their assessment finally become quite bleak.

B. *Irving Fisher*

Beginning in January 1923, Irving Fisher's Business Page included the Fisher Commodity Price Index and the Purchasing Power of the Dollar Index, along with weekly discussions of relevant economic issues of the time. The Commodity Price index is Fisher's "Ideal" index, a geometric average of base and current period weighted indexes. [Fisher, 1923, p. 835] The Purchasing Power of the Dollar index is simply the inverse of the Commodity Price index. The duplication was, however, central to Fisher's missionary ideas for his service. He aimed to educate the public to eliminate money illusion.[7]

In 1925 Fisher added to his analysis a stock market index, which consisted of a weighted value index for the market, with stocks picked based on a standard of "popularity." The index was computed based on averaging the value of the previous week's fifty best-selling stocks, based on total market value of share turnover. The index was linked across weeks by calculating a composite average based on the subsample of stocks that appeared in both weeks' indexes. [Fisher, 1927]

Unlike HES's, Fisher's forecasts were simple and straightforward applications of the classical paradigm, allowing for some rigidities slowing adjustment. Thus, for example, a falling dollar and rising commodity prices "shows that the volume of money and credit available is increasing faster than its utilization in the production and exchange of goods and services. When more money and credit are applied to the same (or smaller) volume of goods and services, it results in increased competition by the funds for the available goods and services."[8] Rising commodity prices would thus engender a forecast of declines in real interest rates, a rising industrial production rate, and increasing inventories due to increases in the value of goods.

---

[6] This decline is not present in either Fisher's stock price index or Alfred Cowles's (1939) stock price index. Hence, it may be spurious.

[7] Fisher analyzes a number of indices published at the time. In a footnote, Fisher comments on the Harvard index methodology: "One of the most interesting kinds of index number is Prof. Person's new index number for use as a barometer of trade. In this case the selection of the 10 commodities included is based, not on any of the usual criteria, but on their previous behavior in relation to the business cycle." [Fisher, 1927, p. 336]

[8] See Edwin Newdick (1929, p. 2). A brief exposition of Fisher's forecasting methodolgy, "How to Use Fisher

Although Fisher's economic analyses are less chartist than HES's, his predictions in the period before and after the Crash were no closer to the mark than those of his Harvard brethren. Indeed, his analyses during 1930 and 1931 were consistently even more optimistic than Harvard's.[9] Immediately after the Crash he noted that price/earnings ratios of less than eleven for industrial stocks were too low given expected earnings and dividend growth in the future. Throughout 1930 Fisher continued to compare favorably the current situation to the 1920–21 recession. Even by late 1931, Fisher continued to read good news into the bad. He believed that commodity prices had temporarily leveled off and business was soon to see a long overdue turnaround.

II. Statistical Analysis

The above discussion of the Harvard and Yale services and the detailed chronology of their pronouncements in the Appendix makes clear that the services neither predicted a downturn before the Crash nor became substantially more pessimistic about the economy following the Crash. In this section, we use time-series methods to study whether these optimistic projections were warranted by the data.

To study the forecastability of the Depression, we estimate a number of vector autoregressive (VAR) models using the Harvard and Fisher data and the modern historical data. The models differ in the number of variables included. The lengths of the estimation periods also differ across models, depending on data availability.

Each equation includes a constant and twelve lags of all the variables in the system. We impose Litterman's Bayesian prior that the variables follow univariate random walks. The standard deviations of the prior take the form

$$(1)\quad S(i,j,k)=\gamma g(k)f(i,j)(s_j/s_i),$$

where i indexes the left-hand side variable, j indexes the right-hand side variables, and k indexes the lag. s_i is the standard error of the unrestricted equation for variable i. We use the parameters $f(i,i)=1.0$, $f(i,j)=.75$, $i\neq j$, $g(k)=k^{-1}$, and $\gamma=0.3$.[10]

We also computed all the estimates with a lag length of two with no priors imposed. These results were quite similar to the results using twelve lags, and so only the forecasts using twelve lags are reported below.

The HES variables are the A, B, and C indexes. The data for these variables begin in January 1919. The Fisher variables are his commodity price index and his stock price index. The data for both of these variables are available beginning in January 1925. The modern historical variables are an industrial production index, a producer price index, a stock price index, a three-month interest rate, and a measure of the money stock. We use two indexes of industrial production, one from the Federal Reserve, which begins in January 1919, and one from Babson, which begins in January 1890.[11] The producer price index is the Bureau of Labor Statistics all-commodity wholesale price index; the stock price index is Cowles all-stock index; and the interest rate is the three-month time loan rate of N. Gregory Mankiw, Jeffrey Miron, and David Weil (1987). The data on these three variables are available beginning at least in January 1890. The money stock

---

Indexes" by Newdick, was part of the *Financial Analysis Service* subscription package.

[9] Again, the chronology in the Appendix supports the following brief summary of Fisher's views.

[10] These priors are somewhat looser than those recommended by Christopher Sims (1987, p. 446). In the RATS programming language, they can be expressed as tightness = 0.3, harmonic decay = 1.0, and other's weight = 0.75. See Thomas Doan and Robert Litterman (1987).

[11] When Babson's index is used, it is scaled to equal the Federal Reserve's index in June 1929. The two indexes are quite close in the period that they overlap. Regressing Babson's index on the Federal Reserve's index and a constant for the period from January 1919 to December 1933 yields an R^2 of .965.

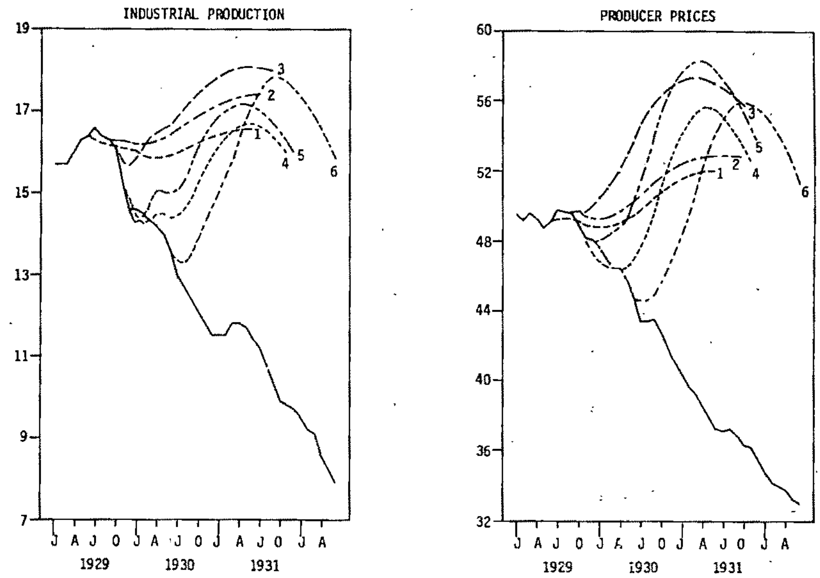


Figure 3. Actual and Forecast Values: Modern Plus HES Data, Sample Begins January 1920. *Note*: System includes industrial production (FRB), producer prices, three-month time loan rate, stock prices, $M2$, and the HES's A, B, and C series. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), December 1929 (5), and June 1930 (6).

series, which begins in May 1907, is Milton Friedman and Anna Schwartz's (1970) $M2$ series.[12] In the estimation work, we use logarithms of the variables except for the A, B, and C variables and the interest rate variable, which were kept in their natural units.

We use the same procedure for each VAR model. The model is first estimated through a particular month. A dynamic, 24-month forecast is then generated using the model, where the forecast begins one month after the end of the estimation period. Six sets of estimates and forecasts are made per model. The first forecast period begins in July 1929 (and so the first estimation period ends in June 1929). The starting dates of the other five forecasts are October 1929, November 1929, December 1929, January 1930, and July 1930. In the figures below we report results only for industrial production and producer prices, although, of course, the forecasts of all the variables in the model underlie these calculations.

The first model consists of eight variables—the five modern variables plus the A, B, and C variables. The estimation periods begin in January 1920. We first test the hypothesis that the A, B, and C variables have coefficients of zero in the industrial production and producer price equations. When two lags are used per variable, the chi-square value is 35.1, which for 12 degrees of freedom rejects the hypothesis at the .01 level. When twelve lags are used, the chi-square value is similar, 38.5, although with the Bayesian priors imposed the number of degrees of freedom is difficult to interpret. At any rate, it thus seems from at least the two-lag case that the A, B, and C variables are jointly significant.

The forecasting results are presented in Figure 3. The solid line indicates the actual data from the beginning of 1929 through the first half of 1932, and the six dashed lines

---

[12] The data and data sources are discussed in the Appendix.

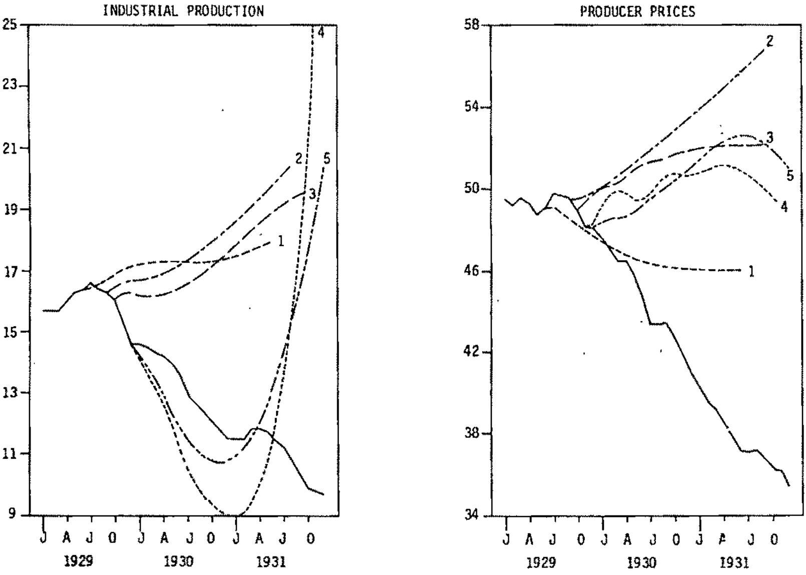


FIGURE 4. ACTUAL AND FORECAST VALUES: MODERN PLUS FISHER DATA, SAMPLE BEGINS JANUARY 1926. *Note*: System includes industrial production (FRB), producer prices, three-month time loan rate, stock prices, $M2$, and Fisher's stock and commodity price indexes. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), and December 1929 (5).

give the 24-month forecasts.[13] It is clear from Figure 3 that neither before nor after the Crash does the VAR system forecast the Depression. Aside from slight dips in output in several of the forecasts for a few months, all the forecasts indicate continued expansion of the economy. All the forecasts thus produce large, negative errors in later periods. The price forecasts similarly systematically overpredict the price level.

The second model consists of the five modern variables plus Fisher's two variables. The estimation periods begin in January 1926. Because Fisher stopped issuing his data in May 1930, we include only the first five forecasts. We also test the hypothesis that the two Fisher variables are zero in the industrial production and producer price equations. In the two-lag case the chi-square value is 14.5, which with 8 degrees of freedom is significant at the .01 level. Again, the value of 14.8 in the twelve-lag case is similar.

Figure 4 presents the forecasting results for this model. The first three forecasts (through the estimation period ending in October 1929) show rapid expansion in output. The implied annual growth rates approach double digits. Once the Crash is embodied in the initial conditions, the forecasts show sharp declines in output followed by sharp rebounds in about a year. Except for the slight deflation forecast before the Crash, the forecasts of the producer price index are essentially flat, and so they imply large forecast errors.

The two models examined so far are based on estimation periods that contain only observations from the 1920s. Aside from the sharp recession at the beginning of the decade, the 1920s witnessed strong and sustained economic growth. It thus may be that the poor performance of the models is simply due to the use of estimation periods that

---

[13] Although the estimates are based on logarithms of industrial production and producer prices, the results are reported for levels of these variables.

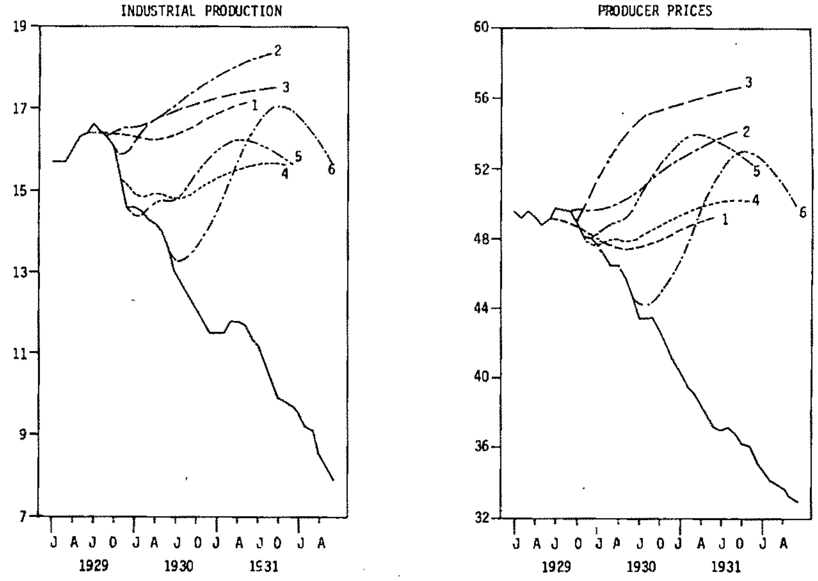


FIGURE 5. ACTUAL AND FORECAST VALUES: MODERN DATA. SAMPLE BEGINS JANUARY 1920. *Note*: System includes industrial production (FRB), producer prices, three-month time loan rate, stock prices, and $M2$. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), December 1929 (5), and June 1930 (6).

are too short. In other words, one might not expect to find seeds of the Depression in the data from the 1920s using time-series methods. We can, however, examine whether the poor forecasting performance of the models is due to the small sample by extending the estimation period back much further. If we drop the HES and Fisher variables and use Babson's index of industrial production in place of the Federal Reserve's, we can extend the sample back to January 1907, and if we further drop the money stock variable, we can extend the sample back to January 1890. Eight troughs occurred between 1890 and 1918 according to the National Bureau of Economic Research chronology (Geoffrey Moore and Victor Zarnowitz, 1986, p. 750), some quite severe. These data thus provide a much firmer basis for attempting to forecast the Depression using time-series methods.

Before extending the sample backward, we present estimates based on the modern variables only, but using the same sample periods as for the HES and Fisher data. These estimates show how the HES and Fisher data affect the forecasts. In Figure 5 we present forecasts from the model using only the five modern variables for the estimation periods beginning in January 1920. These forecasts are comparable to those in Figure 3 except that the HES data are excluded. The forecasts in Figure 5 are close to those in Figure 3, which suggest that the HES variables have little incremental value in forecasting industrial production and producer prices, even though they are jointly significant in the two equations.

We next estimated the model using only the five modern variables for the estimation periods beginning in January 1926. The forecasts are presented in Figure 6. These forecasts are comparable to those in Figure 4 except that Fisher's data are excluded. As in Figure 4, output is forecast to grow much faster than it did, but the size of the forecast errors are smaller than when Fisher's data are included.

Figures 5 and 6 show that the HES and Fisher data do not much affect the forecasts of the Depression based on data from the

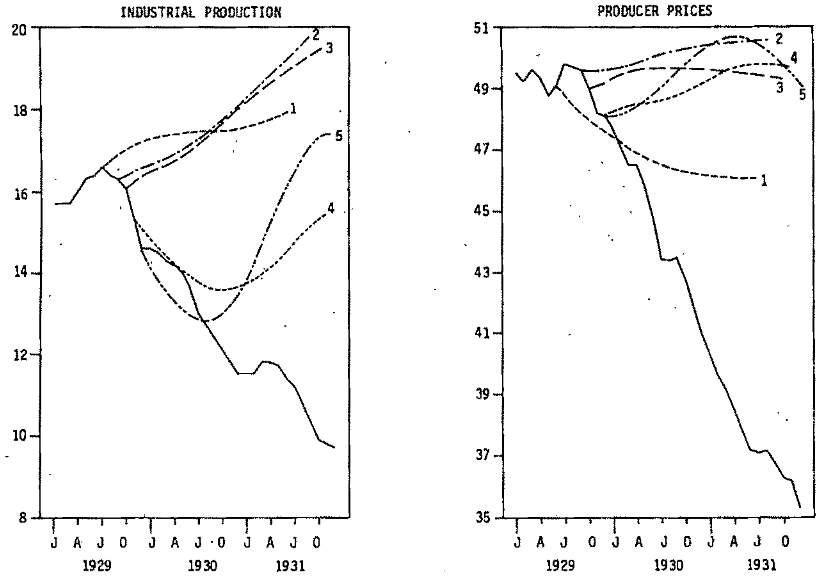


Figure 6. Actual and Forecast Values: Modern Data, Sample Begins January 1926. *Note*: System includes industrial production (FRB), producer prices, three-month time loan rate, stock prices, and $M2$. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), and December 1929 (5).

1920s. We now consider the effects of lengthening the estimation period. The forecasts in Figure 7 are based on the use of the five modern variables for the estimation periods beginning in May 1908.[14] The beginning of the estimation periods is dictated by the money data, which are available beginning in May 1907. The first forecast in Figure 7, based on information through June 1929, shows a moderate downturn in production. This is not true, however, for the other five forecasts. The forecasts for the price level in Figure 7 are more striking. The dramatic deflation of the early 1930s is predicted in the forecasts based on data through the first months after the Crash. Interestingly enough, the last two forecasts, whose initial conditions clearly reflect the actual start of the deflation, do not forecast the continuation of the deflation.

The forecasts in Figure 8 are based on the same estimation periods as those in Figure 7, but the money stock variable is dropped. The model thus consists of only four modern variables. The third forecast in Figure 8 predicts the continuing fall in output fairly well, although the others do not. As in Figure 7, the first three forecasts of the price level predict the deflation quite well. Because similar results are obtained without the money data in Figure 8, the forecasts of the deflation in Figure 7 do not depend on the use of the money data.

Finally, the forecasts in Figure 9 are based on the use of the four modern variables (excluding the money stock) for the estimation periods beginning in January 1891. This sample period is the longest at our disposal with these data. The results show that except for slight downturns in the forecasts based on data through October and November 1929 (the third and fourth forecasts) industrial

---

[14] Given that the estimation period now begins before the Federal Reserve's industrial production index is available, these and subsequent results use the Babson index. Forecasts based on the use of Babson's index for the estimation periods beginning in January 1920 are almost identical to those presented in Figure 5 using the Federal Reserve's index.

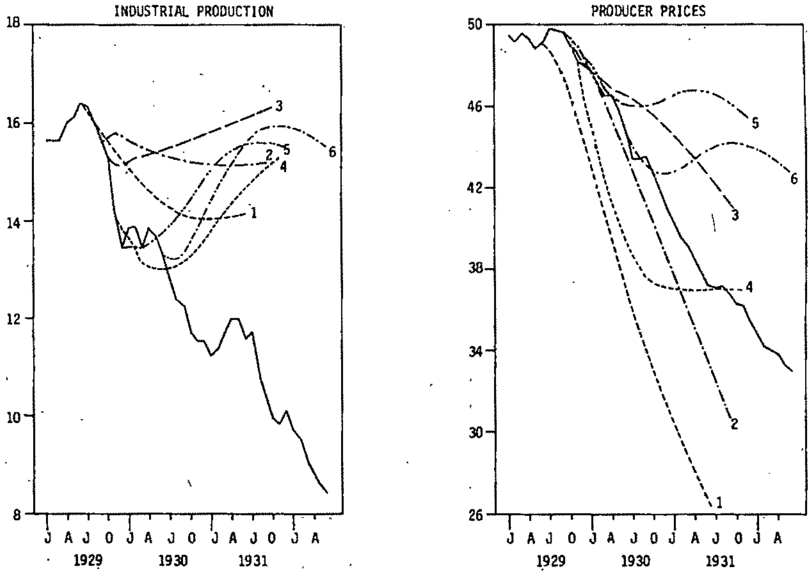


FIGURE 7. ACTUAL AND FORECAST VALUES: MODERN DATA, SAMPLE BEGINS MAY 1908. *Note*: System includes industrial production (Babson), producer prices, three-month time loan rate, stock prices, and $M2$. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), December 1929 (5), and June 1930 (6).

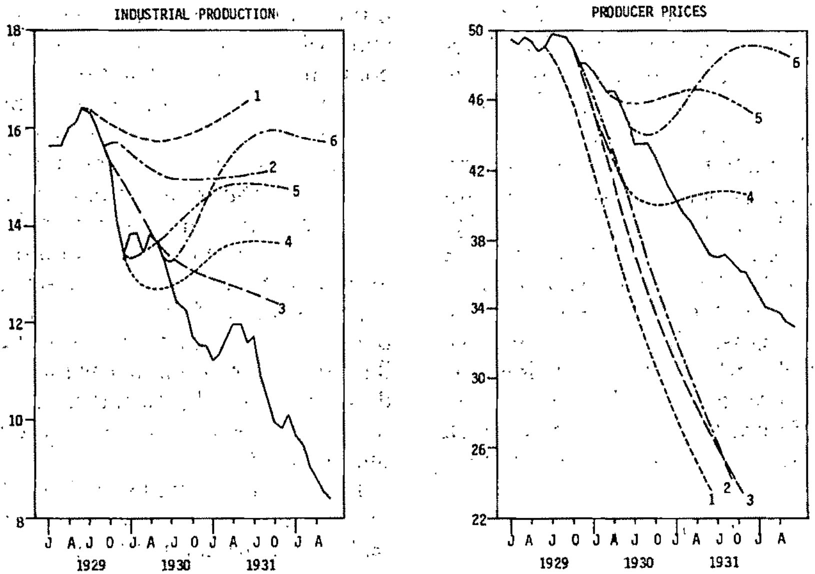


FIGURE 8. ACTUAL AND FORECAST VALUES: MODERN DATA EXCLUDING MONEY STOCK, SAMPLE BEGINS MAY 1908. *Note*: System includes industrial production (Babson), producer prices, three-month time loan rate, and stock prices. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), December 1929 (5), and June 1930 (6).

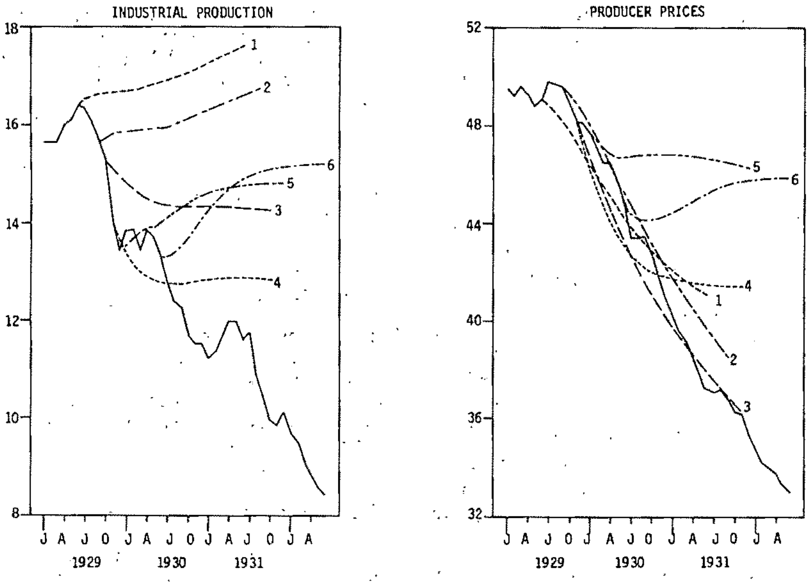


FIGURE 9. ACTUAL AND FORECAST VALUES: MODERN DATA EXCLUDING MONEY STOCK, SAMPLE BEGINS JANUARY 1891. *Note*: System includes industrial production (Babson), producer prices, three-month time loan rate, and stock prices. Solid lines are actual data. Dashed lines are forecasts conditional on data through June 1929 (1), September 1929 (2), October 1929 (3), November 1929 (4), December 1929 (5), and June 1930 (6).

production is predicted to grow on trend. As in Figures 7 and 8, the early forecasts for the price level give very accurate forecasts for the deflation, but these are revised upward in the forecasts based on information through December 1929 and June 1930.

In summary, the results in Figures 3 through 9 show that the large declines in output at the beginning of the Depression were essentially not predicted in both the months before the Crash and once the news of the Crash is taken into account. In a few of the systems, small output declines are predicted, but even these more accurate predictions disappear over longer horizons and as the forecasts adjust to the changing initial conditions. Similarly, the large deflation was not systematically predicted by the systems. In the estimates based on the long samples (Figures 7–9), the early forecasts, which are conditional on data leading up to and just following the Crash, are quite accurate. These forecasting successes should not be overemphasized, however, because the forecasts are very unstable as the initial conditions and coefficient estimates change. Hence, aside from a hint of forecastability of the deflation, the Depression was unforecastable using our time-series methods. Therefore, the Harvard Economic Service, Irving Fisher, or an econometrician endowed with modern time-series methods and data would be justified in appearing optimistic about the economy on the eve of and in the months following the Crash.

III. Conclusion

Our results imply that the Harvard and Yale forecasters cannot necessarily be faulted for remaining optimistic after the Crash. Their continued optimism is consistent with our conclusion based on time-series methods that the Depression was not forecastable. They were both justified in holding what *ex post* was an incorrect view. Thus, one might conclude that Harvard and Yale tied.

Our finding that the Depression was not forecastable should be interpreted cautiously. Our time-series findings do not rule out the possibility that a structural model based on restrictions from economic theory could forecast the Depression. None of the major theories of the Depression, however, necessarily imply that the Depression was forecastable. Friedman and Schwartz (1963) place the blame for the Depression on mistakes made by the Federal Reserve, in particular the fact that the money stock fell by over a third from August 1929 to March 1933. No one suggests that these mistakes were forecastable. Peter Temin (1976) stresses autonomous declines in aggregate demand, and insofar as these were unforecastable, his theory does not make the Depression forecastable. Charles Kindleberger's (1986) account of the Depression stresses the role of financial panics and the collapse of commodity prices. Again, these are not likely to be forecastable. James Hamilton (1987) does argue that the monetary contraction started in 1928. If this early monetary contraction were an important cause of the Depression, our regressions could in principle detect it. Yet, Hamilton (1987, pp. 167–68) does not suggest that the monetary contraction in 1928 was sufficient to cause the Depression, so the failure of time-series techniques to forecast the Depression may not contradict his account. The leading explanations of the Depression are thus based on unforecastable policy and economic disturbances. Therefore, finding that the Depression is unforecastable does not help distinguish among these explanations of its cause.

APPENDIX

Chronology of the Services' Forecasts. This section of the Appendix contains the evolution of pronouncements made by the Harvard Economic Service and Irving Fisher on the status of the economy, from early 1929 through 1931. It clearly demonstrates that both failed to forecast the Depression and remained optimistic following the Crash of the stock market in October 1929.

Harvard Economic Service. In April 1929, HES issued the following warning about the overheated economic environment:

> Recent developments—notably the weakness of certain commodity prices, as well as the money tension—suggest some recession from the present level of business.... A renewed unsettlement of stock prices, perhaps a protracted liquidation, might well result under present conditions of money tension.... Business is not in the strained condition which has led to real depression in the past; and, while interest rates are now higher than at any other time since 1921, the present large resources of the reserve system give assurance that no shortage of credit for industrial purposes will develop.
> [HES, 4/20/1929][15]

This view persisted through the summer of 1929. In May, the Service noted that "The signs pointing to recession continue to pile up" [HES, 5/18/1929].

By mid-September, when the speculation index was rising, the tone moderated considerably:

> Recent developments (reduced volume of construction projects, below average crop prospects, unfavorable international trade balances) have tended to emphasize the unfavorable elements in the business (as distinguished from the financial) situation.... But no sharp decline has appeared in general business, and activity remains high. [HES, 9/21/1929]

The week prior to Black Friday, the assessment was that

> If recession should threaten serious consequences for business (as is not indicated at present) there is little doubt that the reserve system would take steps to ease the money market and so check the movement.
> [HES, 10/19/1929]

This tone persisted as calamities mounted over the coming year. Immediately following the Crash, in early November, the appraisal was that

> Doubtless the losses entailed by the decline in stocks will increase the extent of the present business recession. On the other hand, money rates have declined promptly and considerably... this is in itself evidence of the soundness of the present business situation. Under such conditions, we believe that the present recession, both for stocks and business, is not the precursor of business depression, but will prove intermediate in character. [HES, 11/2/1929]

By late December 1929, the Harvard staff was even forecasting recovery:

> Today a depression seems improbable, and continuance of business recession is all that is in prospect. This justifies a forecast of recovery of business next spring, with further improvement in the fall, so that 1930, as a whole, should prove at least a fairly good year.
> [HES, 12/21/1929]

Confidence then continued as the downturn worsened:

> Since our monetary and credit structure is not only sound but unusually strong, commercial credits are liquid, and production for some months has probably been less than current demands, there is every prospect that the recovery which we have been expecting will not be long delayed, and that the only change that need be made in our forecast is that this fall business will not acquire as much impetus as we have been expecting.
> [HES, 8/20/1930]

---

[15] This and the following quotations from the Harvard Economic Service (HES) appeared in the *Weekly Letter* on the indicated date.

> Except in construction activity, the current depression seems to be following much the same course as that of 1920–21, as respects the length and severity of decline and the inception of recovery; and we conclude that improvement in business volumes, the first step in business recovery, is already underway. [HES, 6/30/1931]

A dawning awareness that this economic downturn was not similar in nature to those preceding it is found in the final months of 1931. In August, the Service noted that "If financial unsettlement continues in Europe, real business recovery here may be prevented..." [HES, 8/22/1931]. And by December of that year, the assessment had finally become quite bleak:

> But though the threatening elements in the situation are fewer, adequate grounds for forecasting business revival have not yet appeared. [HES, 12/19/1931]

Irving Fisher. In September 1929, Fisher was bullish about future economic prospects, seeing none of the potential warning signs that Harvard service had noted through the summer of that year.[16]

> We have witnessed probably the greatest expansion in history, within any similar period of time, of the real income of a people.... The alarm over the supposed inflation of security values seem unjustified by the recent record of dividend yields, increases which doubtless will be followed by much larger increases.
> [Fisher, 9/2/1929][17]

The day before Black Friday, Fisher continued bullish in the face of the (relatively) small perturbations that had rocked the market in recent weeks and stirred concern on Wall Street.

> The [stock market] break certainly exhibited signs of a market rendered topheavy by the activities of shoals of speculators, acting unintelligently, who at last became frightened and dumped millions of shares in a way temporarily to swamp the Exchanges.... The only event which can bring about a serious decline in stock value is a severe business slump, which does not seem likely from present indications. [Fisher, 10/28/1929]

Immediately after the Crash, the outlook remained optimistic:

> The price of industrial stocks at less than 11 times their earnings seems too low a ratio, in view of the expectation of a faster rate of earnings in future and of the diminished risks of modern investment methods.... The market has, therefore, good reason for recovering on a new plateau. [Fisher, 11/4/1929]

As the recession deepened, Fisher's optimism bordered on advocacy, inveighing against dire interpretations of the direction of current indicators:

> It should be noted that while the stock exchange endured the severest panic in history, this did not suffice to shake the price level of stocks off a warrantable high new plateau, which had been built up since 1922.... The function of the stock market is to reflect the discounted value of future increased business, as clearly foreshadowed in the increasing rate at which earnings are plowed-back into industry. [Fisher, 12/30/1929]

> While prediction is always hazardous in economic statistics, and I wish to avoid making any definite prediction, it would not be surprising if by next month the worst of the recession will have been felt and improvement looked for. [Fisher, 1/20/1930]

In the early summer of 1930, Fisher remained unwilling to read calamity into the current economic situation. Comparisons with the 1920–21 recession predominated in his analysis—as they did in Harvard's—but Fisher's brand of optimism remained both stronger and more colorful:

> It seems manifest that thus far the difference between the present comparatively mild business recession and the severe depression of 1920–21 is like that between a thunder-shower and a tornado. [Fisher, 5/19/1930]

By winter of that year, Fisher finally acknowledged the seriousness of the economic situation; yet like members of the Harvard group, he continued to see the potential for an imminent upturn and expressed his hopes with mounting drama:

> Doubtless, to many, this will be the winter of our discontent, and to all of us it is as gloomy as the darkness that preceeds dawn.... One would expect the bottom to be reached in the stock market prior to the upturn in commodity prices. Since the middle of November this seems to have been the case.
> [Fisher, 12/1/1930]

> In attempting to review and estimate the business situation, it is sometimes necessary, during a business depression such as this, to find hopeful signs in negative figures. Such a potential state of affairs holds true just now. [Fisher, 2/2/1931]

> The industrial giant has become conscious again. He is beginning to move around slowly in an effort to regain his feet, as the cobwebs from a knockout blow gradually clear from his brain. [Fisher, 2/23/1931]

By the end of 1931, when the Harvard team had begun to admit to the reality contained in the steady flow of bleak economic data, Fisher maintained his optimism.

> Business showed further gain last week, and if improvement continues at the present rate, September should mark the low of the depression. Particular attention is called to the fact that commodity prices for both raw materials and agricultural products were strong. No single factor would be as helpful to business as a cessation in the commodity price decline that has been continued for the past two years. [Fisher, 10/17/1931]

---

[16] Fisher's optimism, however, is consistent with his data. In contrast to the Harvard speculation index (A), which shows a 10 percent decline from March to June 1929, the Fisher stock price index shows no decline. Similarly, the Cowles (1939) stock price index, which has been widely used in recent studies, has some fluctuations but is essentially flat over the interval.

[17] This and subsequent statements appeared in Fisher's syndicated Business Page on the indicated dates. Copies of the columns are contained in the Fisher collection housed in the Manuscript and Archive Collection, Sterling Memorial Library, Yale University.

Data Sources. This section gives the details of the data used in this paper. Table A1 gives the HES's *A*, *B*, and *C* series and Fisher's stock and commodity prices. The source for these series and the modern historical data are as follows:

Harvard Economic Services Curve-*A*, Curve-*B*, and Curve-*C* were collected from various issues of *The Review of Economic Statistics* 1919–22, and the Harvard *Weekly Letter*, 1923–31.

Fisher's Wholesale Price Index appeared in various issues of the *Journal of the American Statistical Association*, 1923–30; a revised series appeared in *JASA* Reprint, September 1930.

Fisher's Stock Index was collected from Fisher's Scrapbooks 1925–27 (Yale Manuscript and Archive Collection) and *Index Number Institute* releases 1928–30.

The Industrial Production Index is a revised seasonally adjusted series published by the *Federal Reserve Board* in Statistical Release G12.3, July 18, 1985.

Babson's index of industrial production closely tracks the FRB's during the period of overlap. It is published beginning monthly on a seasonally adjusted basis beginning in 1889 in Moore (1961, p. 130).

The interest rate series is the three-month time loan rate collected and analyzed by Mankiw, Miron, and Weil (1987, p. 372). These data are available beginning in 1890.

The money series is Friedman and Schwartz's (1970) *M2* series. They are available on a monthly seasonally adjusted basis beginning in May 1907.

Our price index is the Bureau of Labor Statistics Wholesale Price Index (now called the Producer Price Index). The annual wholesale price index is Series E23 in *Historical Statistics*. We obtained the monthly series from the BLS beginning in 1913. The monthly series from 1890 through 1912 is from Bureau of Labor Statistics, *Index Number of Wholesale Prices on Pre-War Base, 1890–1927* (1928). We have been unable to locate this series in present-day BLS sources.

The stock price is Cowles's (1939) all commodity stock price index (Series *P*-1).

TABLE A1—HARVARD ECONOMIC SERVICE AND IRVING FISHER'S DATA

Month and Year	Harvard Economic Service			Fisher	
	Speculation (A)	Business (B)	Money (C)	Commodity Prices	Stock Prices
Jan. 19	−1.13	−0.88	0.39	NA	NA
Feb. 19	−1.00	−1.14	0.36	NA	NA
Mar. 19	−0.96	−1.46	0.45	NA	NA
Apr. 19	−0.44	−1.23	0.55	NA	NA
May 19	0.62	−0.40	0.56	NA	NA
June 19	1.00	−0.16	0.73	NA	NA
July 19	1.69	0.35	0.74	NA	NA
Aug. 19	1.22	0.54	0.64	NA	NA
Sept. 19	1.24	0.18	0.52	NA	NA
Oct. 19	1.63	0.26	0.61	NA	NA
Nov. 19	1.62	0.34	0.83	NA	NA
Dec. 19	1.16	0.76	1.02	NA	NA
Jan. 20	0.88	1.07	1.46	NA	NA
Feb. 20	0.38	0.95	1.87	NA	NA
Mar. 20	0.82	1.20	1.91	NA	NA
Apr. 20	0.88	1.44	1.93	NA	NA
May 20	0.15	1.76	2.27	NA	NA
June 20	0.05	1.94	2.44	NA	NA
July 20	0.14	2.02	2.50	NA	NA
Aug. 20	−0.19	1.68	2.61	NA	NA
Sept. 20	−0.28	1.50	2.33	NA	NA
Oct. 20	−0.36	0.84	2.20	NA	NA
Nov. 20	−0.34	0.35	2.17	NA	NA
Dec. 20	−0.64	−0.01	1.95	NA	NA
Jan. 21	−0.97	−0.58	1.85	NA	NA
Feb. 21	−1.10	−0.88	1.87	NA	NA
Mar. 21	−1.24	−1.27	1.73	NA	NA
Apr. 21	−1.20	−1.25	1.68	NA	NA
May 21	−0.95	−1.09	1.57	NA	NA
June 21	−1.30	−1.14	1.53	NA	NA
July 21	−1.36	−1.42	1.10	NA	NA
Aug. 21	−1.46	−1.16	0.85	NA	NA

TABLE A1—CONTINUED

	Harvard Economic Service			Fisher	
Month and Year	Speculation (A)	Business (B)	Money (C)	Commodity Prices	Stock Prices
Sept. 21	−1.31	−1.08	0.63	NA	NA
Oct. 21	−1.41	−1.09	0.46	NA	NA
Nov. 21	−1.04	−1.35	0.24	NA	NA
Dec. 21	−0.74	−1.44	0.20	NA	NA
Jan. 22	−0.94	−1.66	0.07	NA	NA
Feb. 22	−0.74	−1.34	0.07	NA	NA
Mar. 22	−0.60	−1.25	−0.06	NA	NA
Apr. 22	−0.24	−1.20	−0.26	NA	NA
May 22	−0.14	−0.64	−0.35	NA	NA
June 22	−0.20	−0.38	−0.46	NA	NA
July 22	−0.02	−0.38	−0.55	NA	NA
Aug. 22	0.20	−0.08	−0.61	NA	NA
Sept. 22	0.28	−0.14	−0.39	NA	NA
Oct. 22	0.30	−0.10	−0.22	NA	NA
Nov. 22	−0.02	−0.34	−0.06	NA	NA
Dec. 22	−0.12	0.10	−0.06	NA	NA
Jan. 23	−0.04	0.52	−0.14	100.6	NA
Feb. 23	0.14	0.71	−0.01	103.2	NA
Mar. 23	0.24	0.69	0.24	106.7	NA
Apr. 23	0.12	0.72	0.27	107.6	NA
May 23	−0.12	0.98	0.31	104.5	NA
June 23	−0.17	0.62	0.14	101.2	NA
July 23	−0.43	0.20	0.22	97.0	NA
Aug. 23	−0.48	0.38	0.22	96.9	NA
Sept. 23	−0.49	−0.13	0.29	98.8	NA
Oct. 23	−0.56	−0.06	0.20	100.0	NA
Nov. 23	−0.34	−0.03	0.14	98.4	NA
Dec. 23	−0.24	−0.16	0.07	97.8	NA
Jan. 24	−0.16	−0.14	0.09	98.0	NA
Feb. 24	−0.09	0.39	0.00	99.9	NA
Mar. 24	−0.10	0.01	−0.12	98.0	NA
Apr. 24	−0.26	0.10	−0.25	96.6	NA
May 24	−0.33	0.24	−0.45	96.0	NA
June 24	−0.18	−0.12	−0.76	95.4	NA
July 24	0.03	0.16	−1.14	97.3	NA
Aug. 24	0.21	0.36	−1.29	98.9	NA
Sept. 24	0.09	−0.01	−1.32	99.2	NA
Oct. 24	0.08	0.10	−1.35	101.4	NA
Nov. 24	0.62	−0.08	−1.20	102.6	NA
Dec. 24	0.93	0.24	−0.92	103.5	NA
Jan. 25	0.84	0.66	−0.81	106.7	74.6
Feb. 25	0.90	0.66	−0.79	107.5	76.5
Mar. 25	0.74	0.58	−0.59	107.3	74.7
Apr. 25	0.50	0.66	−0.63	103.5	73.0
May 25	0.79	0.56	−0.67	103.5	77.7
June 25	0.82	0.64	−0.63	104.0	82.1
July 25	0.94	0.78	−0.60	105.8	85.3
Aug. 25	1.26	0.50	−0.49	105.6	88.5
Sept. 25	1.26	0.69	−0.37	104.6	93.5
Oct. 25	1.50	1.06	−0.21	103.6	102.4
Nov. 25	1.78	0.64	−0.17	105.2	106.8
Dec. 25	1.87	0.75	−0.12	105.2	107.8
Jan. 26	1.82	1.65	−0.14	105.4	111.2
Feb. 26	1.74	1.28	−0.24	104.2	114.4
Mar. 26	1.35	1.11	−0.27	101.0	101.2
Apr. 26	1.24	1.06	−0.47	99.6	94.9
May 26	1.34	1.01	−0.49	100.5	95.5

TABLE A1—CONTINUED

Month and Year	Harvard Economic Service			Fisher	
	Speculation (A)	Business (B)	Money (C)	Commodity Prices	Stock Prices
June 26	1.75	0.98	−0.50	100.5	105.0
July 26	2.04	1.14	−0.43	98.8	114.1
Aug. 26	2.40	0.84	−0.27	97.5	127.2
Sept. 26	2.46	0.74	−0.18	97.9	127.2
Oct. 26	2.18	0.99	−0.14	98.0	117.9
Nov. 26	2.19	0.55	−0.23	99.7	119.4
Dec. 26	2.47	0.65	−0.23	97.3	126.3
Jan. 27	2.40	0.82	−0.28	96.0	125.1
Feb. 27	3.06	0.76	−0.41	94.4	131.4
Mar. 27	3.27	0.72	−0.46	92.9	139.9
Apr. 27	3.63	0.76	−0.45	92.5	147.4
May 27	3.82	0.80	−0.35	92.5	156.0
June 27	4.08	0.66	−0.30	92.4	164.9
July 27	4.14	0.78	−0.32	92.2	174.1
Aug. 27	4.46	0.66	−0.55	93.2	186.7
Sept. 27	4.88	1.05	−0.65	95.8	207.2
Oct. 27	4.66	1.20	−0.63	96.2	209.4
Nov. 27	4.56	0.98	−0.62	96.4	216.9
Dec. 27	5.11	1.22	−0.63	96.1	245.3
Jan. 28	5.32	1.23	−0.47	95.5	260.0
Feb. 28	5.07	0.79	−0.41	96.7	263.0
Mar. 28	5.52	1.26	−0.41	97.6	295.4
Apr. 28	6.26	1.82	−0.21	99.1	339.3
May 28	6.94	1.93	0.07	99.1	369.1
June 28	6.37	1.92	0.43	98.0	358.0
July 28	6.23	1.38	0.56	99.6	365.0
Aug. 28	3.70	1.16	0.71	99.8	399.4
Sept. 28	4.72	1.46	0.86	99.7	470.4
Oct. 28	5.54	1.33	0.92	98.8	527.4
Nov. 28	6.82	1.20	0.92	97.5	606.3
Dec. 28	7.14	1.55	1.15	97.0	636.4
Jan. 29	8.28	1.40	1.33	97.4	717.0
Feb. 29	8.48	1.31	1.33	97.7	749.5
Mar. 29	8.98	1.64	1.43	98.3	823.8
Apr. 29	8.44	1.54	1.75	97.1	837.9
May 29	8.30	1.26	1.95	95.9	874.6
June 29	8.08	1.38	1.85	96.9	844.6
July 29	9.27	2.03	1.68	98.5	951.3
Aug. 29	10.08	1.96	1.92	97.3	1038.3
Sept. 29	10.63	1.74	1.90	96.0	1132.1
Oct. 29	8.33	1.84	1.43	94.4	998.1
Nov. 29	4.60	1.30	0.49	92.7	719.7
Dec. 29	4.68	0.58	0.05	92.8	735.8
Jan. 30	4.80	0.31	0.04	93.3	731.6
Feb. 30	5.78	0.16	−0.08	92.7	808.2
Mar. 30	6.48	0.14	−0.51	90.8	885.1
Apr. 30	7.29	0.10	−0.63	90.6	985.2
May 30	6.31	0.02	−0.78	88.6	921.5
June 30	4.61	−0.33	−1.03	86.4	NA
July 30	4.26	−1.04	−1.27	NA	NA
Aug. 30	3.96	−1.23	−1.41	NA	NA
Sept. 30	3.84	−1.32	−1.47	NA	NA
Oct. 30	2.26	−1.74	−1.53	NA	NA
Nov. 30	1.68	−2.24	−1.66	NA	NA
Dec. 30	1.16	−2.50	−1.59	NA	NA
Jan. 31	1.35	−2.63	−1.73	NA	NA
Feb. 31	2.25	−2.90	−1.86	NA	NA

TABLE A1—CONTINUED

Month and Year	Harvard Economic Service			Fisher	
	Speculation (A)	Business (B)	Money (C)	Commodity Prices	Stock Prices
Mar. 31	2.10	−3.25	−1.85	NA	NA
Apr. 31	1.42	−3.17	−1.93	NA	NA
May 31	0.39	−3.45	−2.96	NA	NA
June 31	0.39	−3.72	−3.22	NA	NA
July 31	0.44	−3.96	−3.20	NA	NA
Aug. 31	−0.08	−4.14	−3.23	NA	NA
Sept. 31	−0.72	−4.41	−3.12	NA	NA
Oct. 31	−1.16	−4.84	−1.87	NA	NA
Nov. 31	−1.42	NA	−1.26	NA	NA

REFERENCES

Cowles, Alfred 3rd and Associates, *Common Stock Indexes*, Cowles Commission for Research in Economics Monograph No. 3, 2nd ed., Bloomington, IN: Principia Press, 1939.

Doan Thomas, A. and Litterman, Robert B., *RATS User Manual*, VAR Econometrics, Evanston, IL, 1987.

Fisher, Irving, *The Making of Index Numbers*, New York: Houghton-Mifflin, 1922.

______, "A Weekly Index Number of Wholesale Prices," *Journal of the American Statistical Association*, September 1923, *18*, 835–40.

______, "Wholesale Price Index of Commodities, 1923–30," *Journal of the American Statistical Association*, reprint, September 1930.

______, "The 'Total Value Criterion': A New Principle in Index Number Construction," *Journal of the American Statistical Association*, December 1927, *22*, 419–41.

Friedman, Milton and Schwartz, Anna J., *A Monetary History of the United States 1867–1960*, Princeton, NJ: Princeton University Press, 1963.

______ **and** ______, *Monetary Statistics of the United States*, New York: NBER, 1970.

Hamilton, James, "Monetary Factors in the Great Depression," *Journal of Monetary Economics*, March 1987, *19*, 145–69.

Kindleberger, Charles P., *The World In Depression, 1929–1939*, revised ed., Berkley, CA: University of California Press, 1986.

Mankiw, N. Gregory, Miron, Jeffrey A. and Weil, David N., "The Adjustment of Expectations to a Change in Regime: A Study of the Founding of the Federal Reserve," *American Economic Review*, June 1987, *77*, 358–74.

Moore, Geoffrey H., ed., *Business Cycle Indicators, Basic Data on Cyclical Indicators*, Vol. II, Princeton: Princeton University Press, 1961.

______, **and Zarnowitz, Victor,** "The Development and Role of the National Bureau of Economic Research's Business Cycle Chronologies," in Robert J. Gordon, ed., *The American Business Cycle: Continuity and Change*, Chicago: University of Chicago Press, 1986, Appendix A.

Newdick, Edwin, *How to Use Fisher Indexes*, New York: Index Number Institute, 1929.

Persons, Warren, *Interpretation of the Index of General Business Conditions*, Cambridge, MA: Harvard University Press, 1922.

______, "The Revised Index of General Business Conditions," *Review of Economic Statistics*, July 1923, *5*, 187–95.

Sims, Christopher A., "Comment," *Journal of Business and Economic Statistics*, October 1987, *5*, 443–49.

Temin, Peter, *Did Monetary Forces Cause the Great Depression?*, New York: Norton, 1976.

Harvard Economic Society, *Weekly Letters*, Harvard University Press: Cambridge, MA, *1–10*, 1923–31.

Harvard University Committee on Economic Research, (1923a) *Harvard Economic Service*, Cambridge, MA: Harvard University

Press, 1923.

______, (1923b) *The Harvard Index of General Business Conditions: Its Interpretation*, Cambridge, MA: Harvard University Press, 1923.

______, *The Review of Economic Statistics*, Cambridge, MA: Harvard University Press, Preliminary Volumes, *1–4*, 1919–22.

Index Number Institute, *Statistical Release*, Index Number Institute, Inc., New York, 1928–30.

U.S. Bureau of the Census, *Historical Statistics of the United States, Colonial Times to 1970*, Washington: USGPO, 1975.

U.S. Bureau of Labor Statistics, *Index Number of Wholesale Prices on the Pre-War Base, 1890–1927*, Washington: USGPO, 1928.

Unemployment in the Soviet Union: Evidence from the Soviet Interview Project

By PAUL R. GREGORY AND IRWIN L. COLLIER, JR.*

Unemployment statistics are not published by the Soviet government which claims unemployment was "liquidated" in the early 1930s. An unemployment rate for the Soviet urban population during the late 1970s is calculated from a survey of 2793 former Soviet citizens. Unemployment incidence, multiple spells, and unemployment duration by demographic characteristics are compared with U.S. patterns. Similar unemployment rates for Soviet men and women and a positive relation between education and unemployment are found.

In this paper we examine the unemployment experienced by former Soviet citizens before their emigration to the United States between January 1979 and April 1982. In 1983, the National Opinion Research Center administered an interdisciplinary questionnaire prepared by the Soviet Interview Project to 2793 Soviet immigrants living in the United States.[1] The sample was drawn from more than 33,000 Soviet immigrants using selected ethnic, educational, and geographic stratification characteristics of the referent population—the adult Soviet European population residing in medium to large cities.[2] About 90 percent of respondents were Jewish. The respondents reported on their political, social, cultural, and economic activities during the last five years of normal life in the Soviet Union before the emigration decision. For most respondents, this last normal period occurred sometime during 1974–1980.[3]

I. Unemployment in the Soviet Economy

Unemployment statistics are not published by the Soviet government which claims unemployment was "liquidated" in the early 1930s.[4] However knowledgeable Western observers agree that the Soviet labor market shares many of the characteristics of Western labor markets, including unemployment.[5] While official Soviet statistics do not publish such data, specialized Soviet monographs report data on turnover rates, job separations, quits, firings, and duration between jobs.[6] Thus at any time there are Soviet workers who would be counted as unemployed

---

*Department of Economics, University of Houston, Houston, TX 77004. Data for this study were produced by the Soviet Interview Project supported by Contract No. 701 from the National Council for Soviet and East European Research to the University of Illinois, Urbana-Champaign, James R. Millar, principal investigator. The analysis and interpretations of this paper are those of the authors, not necessarily of the sponsors. We are grateful to David Granick and Thomas Mayor for helpful comments on an earlier version of the paper. We also wish to thank two anonymous referees for their comments and suggestions.

[1] See National Opinion Research Center, 1985, for a detailed description of procedures. Copies of the questionnaire, methodological instructions, codebooks, and the Soviet Interview Project dataset are available from the Inter-University Consortium for Political and Social Research, PO Box 1248, Ann Arbor, MI 48106-1248.

[2] The Soviet Interview Project General Survey sample is discussed in Barbara Anderson and Brian Silver, 1987a. The Soviet referent population is analyzed by Anderson and Silver, 1984, and Anderson, Silver, and Robert Lewis, 1986. The issue of response effects is considered in Michael Swafford et al., 1987.

[3] For 89 percent of the respondents, the end of the last normal period occurred during the period 1978–80. The mode of the distribution of end points of the last normal period occurs in 1979 (42 percent), hence our identification of the years 1974–79 as the approximate reference period for the unemployment rates we have calculated.

[4] See R. W. Davies, 1986; Davies and S. G. Wheatcroft, 1986; and Judith Shapiro, 1987.

[5] See, for example, David Granick, 1987a, b; Abram Bergson, 1964, ch. 6; and Gur Ofer and Aron Vinokur, 1981.

[6] These surveys are summarized in Granick, 1987a and A. E. Kotliar, ed., 1983, for example, pp. 57–8.

according to the official U.S. definition—not employed, actively looking for a job, and currently available for work.

Institutional peculiarities of Soviet labor markets could affect the extent and the pattern of unemployment. Soviet citizens are subject to antiparasitism laws. With very few exceptions jobless workers do not receive unemployment compensation.[7] Privileges, not otherwise available through (legal) market transactions, can account for a large proportion of compensation. College graduates are administratively assigned to their first jobs. Soviet enterprises operate in a "job rights" economy in which workers acquire a form of job tenure through an implicit social contract.[8] Even in the absence of job rights, Soviet managers would be reluctant to dismiss low productivity workers since enterprise performance is judged on plan fulfillment (meaning in the first instance achieving quantity targets) and Soviet enterprise managers operate in tight labor markets.

Soviet labor economists characterize spells of joblessness as relatively brief, voluntary interruptions of employment. Should the spell become protracted, it is presumed that the worker is unwilling to accept a reasonable job offer.[9] In fact, the zero percent unemployment rates officially claimed by the Soviet government are the consequence of refusing to count individuals as unemployed who might have or are presumed to have job offers.[10]

II. Measuring Soviet Unemployment from the Soviet Interview Project Sample

As a one-time retrospective survey, the Soviet Interview Project gathered unemployment information differently from Western unemployment surveys. The first difference is the relatively long reference period of five years—the last five years of normal life in the Soviet Union. The U.S. *Current Population Survey* (CPS), in contrast, uses a one-week reference period. For a sample of less than three thousand respondents with a low incidence of unemployment, the longer reference period is required to collect a sufficient number of spells for analysis.

A second difference is the retrospective nature of the Soviet Interview Project survey. The CPS targets current labor force activity. Closer to the spirit of the Soviet Interview Project is the annual U.S. *Work Experience Survey* (WES) which asks about labor force activity during the preceding year.[11] Some Soviet Interview Project respondents were expected to recall events that took place a decade or more before the interview whereas WES respondents are asked to recall experiences only two to fourteen months after the event. The Soviet Interview Project's five-year reference period raises the question of recall bias which is discussed below.

The final important difference is that the Soviet respondents were asked about only those unemployment spells of one month or longer duration.[12] The unemployment data for both the CPS and WES include all unemployment spells. The truncation at one-month duration was selected for two reasons: 1) one month corresponds roughly to the dividing line between short- and long-run unemployment in U.S. statistics (five weeks), and 2) a monthly partition was expected to capture "salient" spells of unemployment and minimize recall problems (compare the discussion of the work of George Akerlof and Janet Yellen below.)

Unemployment histories of Soviet Interview Project respondents were gathered through a series of three questions:

---

[7] For minor exceptions, see Granick, 1987b.

[8] The institutional peculiarities of the Soviet labor market are discussed in Philip Hanson, 1986, and Nick Lampert, 1986. The job rights terminology is developed in Granick (1987a).

[9] To quote from a leading Soviet work on this subject: "Under conditions of socialism, the condition of being without work (*nezaniatost'*) is not a synonym for 'unemployment.' It means only an interruption of work caused by reasons of a private character (family circumstances, changes of location)." Kotliar, 1983, p. 9.

[10] Davies and Wheatcroft, 1986, pp. 36–49.

[11] For 1969 and 1976, see U.S. Department of Labor, 1971, 1977.

[12] For the remainder of the paper all mention of Soviet unemployment rates, incidence, spells, and duration refers exclusively to one-month-or-more spells of unemployment.

> During the 5 years before the end of the last normal period,[13] was there any time that you were unemployed when you looked for a job but couldn't find a job for at least 30 days?
> Yes or No.

For those answering yes, two follow-up questions were asked:

> How many times were there during those five years?
> Number of instances.

The last question of the series was:

> When was that time (that/the last such) time before the end of the last normal period?
> From: month and year.
> To: month and year.

From responses to these three questions, the unemployment rate of the Soviet Interview Project sample was calculated according to the following formula:

$$U=(t\times N\times AD)/60, \quad (1)$$

where U is the average unemployment rate for a five-year period;
t is the percent of respondents not out of the labor force for reasons of retirement, health, or full-time study who report at least one unemployment spell;
N is the average number of unemployment spells for those respondents reporting spells;
AD is the average duration of unemployment spells;
60 is the number of months in a five-year period.

The expression in parentheses is the average number of months in a five-year period that an average respondent in the labor force was unemployed.[14]

III. Reliability, Biases, and Adjustments

A spell of one-month-or-more unemployment over the five years of the last normal period is a relatively rare event among Soviet Interview Project respondents. We examined each case of reported unemployment for reliability, consistency, and plausibility. After the raw data were checked to remove incorrect or inconsistent responses,[15] only 160 of the 2793 respondents reported any spells during the five-year period. The process of sifting through the individual cases revealed many of the recall and reporting problems typical of retrospective data.[16] Moreover, minor adjustments had to be made for unemployment spells that overlapped either the beginning or the end of the five-year period.

The removal of cases of over-reporting eliminates most of the upward bias in the reported number of unemployment spells and their duration. However, significant uncorrected biases remain which run in the opposite direction. It is not possible to detect respondents who failed to report unemployment spells either due to recall problems, a misunderstanding of the question, or other reasons.[17] Because of this asymmetry between our ability to adjust for over-reporting and to detect underreporting, we believe that unemployment spells for the Soviet Interview Project sample are underreported. *Therefore unemployment rates calculated from Soviet Interview Project data displayed in Tables 1 and 4 represent lower bounds for the true unemployment rates.*

---

[13] The end of the last normal period was established by answers to earlier questions: 1. "When did you first apply for an exit visa?"; 2. "Because of your plans did your life change significantly before [the date of your application]?"; if yes, 3. "What was that date?" The end of the last normal period for each respondent was defined as the month preceding the date given in response to the last question. If no significant change had occurred in the respondent's life as a consequence of planning to apply for an exit visa, then the month preceding the application date (the first question) was defined as the end of the last normal period.

[14] Parallel to U.S. labor force statistics, we have defined the state of being in the labor force as either being employed or unemployed.

[15] These adjustments are described in the Appendix.

[16] For example, several respondents mistakenly reported unemployment spells which had occurred before the last five years of normal life in the Soviet Union. The tendency to overstate the total number of events occurring during a time interval (sometimes out of a desire to "please" the interviewer) is called "telescoping." See the discussion in Akerlof and Yellen, 1985, pp. 757–8. A few respondents apparently confused the state of not being in the labor force with unemployment—for example, one woman in her sixties who claimed 60 months of unemployment during the five years of the last normal period. The interested reader is referred to the Technical Appendix of our Soviet Interview Project working paper, Paul Gregory and Irwin Collier, 1987.

P 4419

Comparisons of CPS and WES unemployment rates provide an approximate lower bound on the amount of retrospective underreporting of unemployment. U.S. respondents tend to filter out "nonsalient" unemployment spells, causing unemployment rates calculated from retrospective data to understate unemployment.[18] Akerlof and Yellen, 1985, found unemployment rates from the retrospective WES 10 percent below the rates from contemporaneous surveys. Francis Horvath's, 1982, study for the decade 1969 to 1979 yielded a 19 percent understatement for retrospective unemployment surveys.[19]

Soviet Interview Project respondents had to define for themselves the state of unemployment and could have easily failed to recognize spells associated with entry or reentry into the labor force or finding a job after completing formal education. Because the interviews were conducted with a longer time lapse between the reference period and interview date and covered a longer reference period, we expect that Soviet Interview Project underreporting of unemployment spells would be greater than the 10 to 19 percent underreporting found in U.S. retrospective surveys.[20]

IV. Soviet Interview Project Unemployment Rates and Patterns

Table 1 reports sample frequencies of unemployment spells, average number of spells, average duration of unemployment, and unemployment rates for the Soviet sample and for demographic subgroups. The most important assumptions underlying the calculated unemployment rates (described in the Appendix) are that the labor force participation rate at the end of the last normal period is representative of the whole five-year period and that the duration of the last spell is typical of previous spells for the relatively few respondents with more than one spell.

Table 1 shows that 8.2 percent of respondents in the labor force experienced spells of unemployment during the five-year period. Of those who experienced unemployment, the average number of spells was 1.6, and the average duration of a spell was 4.8 months. Using equation (1), the average respondent in the labor force was unemployed .63 months over a five-year period for an unemployment rate of 1.1 percent. If we apply the U.S. correction factors for retrospective data (10 to 19 percent), the estimated unemployment rate would increase to 1.2–1.3 percent. For reasons mentioned above, we regard the 1.2–1.3 percent figure as a reasonable lower bound for the "true" Soviet Interview Project unemployment rate.

We see from Table 1 that Soviet Interview Project unemployment rates are *inversely* associated with age and *positively* related to the level of education. They do not appear to be strongly affected by the respondent's sex. Equation (1) allows us to decompose the

---

[17] Survey research generally finds that reported occurrence rates of past events decline as the reference period is lengthened. See Seymour Sudman and Norman Bradburn, 1973, cited in Akerlof and Yellen, 1985, p. 756. With a five-year recall period, the degree of understatement of unemployment spells by Soviet Interview Project respondents could be quite high. Research by Anderson and Silver, 1987b, on responses by pairs from the same household reveals that Soviet Interview Project respondents have excellent recall of objective factual information. However, they find less agreement on questions where concepts are complex. The accurate recall of objective factual data, while encouraging, does not establish that the recall of unemployment spells will be equally accurate.

[18] Akerlof and Yellen, 1985, and Francis Horvath, 1982.

[19] Horvath, 1982, p. 41.

[20] Nonetheless in Table 5 below we simply use the 10–19 percent adjustment for memory decay.

TABLE 1—ONE-MONTH-OR-MORE UNEMPLOYMENT RATES BY GENDER, AGE, AND EDUCATIONAL ATTAINMENT, 1974–79

	Soviet Interview Project				
	Respondents Reporting Unemployment[a] Number	Percent of Labor Force with at Least One Spell of Unemployment in last 5 Years	Average Number of Spells for Respondents Reporting Unemployment (Spells/5 years)	Average Duration of a Spell (Months/Spell)	Estimated Unemployment Rate (in percent)
All Respondents	160	8.2	1.6	4.8	1.1
By Gender					
Men	84	8.8	1.8	4.4	1.1
Women	76	8.0	1.4	5.4	1.0
By Age at Last Spell					
18–24	28	26.8	1.3	6.0	3.5
25–34	76	11.5	1.8	4.1	1.5
35–49	36	4.7	1.4	4.3	0.5
50+	20	2.6	2.2	4.0	0.4
By Educational Attainment					
Attestat or Less	23	4.7	1.8	4.6	0.7
Some Secondary	45	8.0	1.6	4.3	0.9
Any University	92	10.8	1.6	5.1	1.5

[a]Age of respondent at the end of last normal period.

unemployment rates into incidence and duration factors: The higher unemployment rates of young respondents are due to a much higher frequency of unemployment spells (26.8 percent of ages 18–24 had at least one spell versus 4.7 percent of ages 35–49) and to longer-duration spells (6.0 versus 4.3 months). These two effects are offset somewhat by a lower proportion having multiple spells. The higher unemployment rates of the university educated are the result of a higher incidence of spells (10.8 percent for university-educated versus 4.7 percent for high-school graduates or less) and a slightly higher average duration of spells (5.1 versus 4.6 months). The slightly lower incidence of spells for women (8.0 percent versus 8.8 percent for men) offsets their longer duration (5.4 versus 4.4 months for men) to yield similar unemployment rates for men and women.

The one-way relationships between unemployment and any single demographic factor seen in Table 1 could be misleading since other demographic factors are not held constant. Regressions of unemployment incidence, number of spells of those reporting spells, and duration of last completed spell on demographic characteristics are reported in Table 2.[21]

Sample selection for the probit regression (column A) was based upon the Soviet Interview Project's Life History Chart described in the Appendix. If a respondent worked for

---

[21] In our decomposition of unemployment rates according to the factors of equation (1), we assumed that the labor force participation rates at the end of the last normal period (for which we have very good data) are representative of the entire five-year period. This procedure is appropriate for computing the averages in Table 1 but not the same as the sample selection procedure for the incidence probit regressions described in our notes to Table 2. The careful reader is assured that the apparent discrepancy between the Table 1 figure, 8.2 percent of the labor force reporting at least one one-month-or-more spell, and the implicit proportion of 6.2 percent (obtained by dividing the 160 cases reported in Table 1 by the sample size of 2331 at the bottom of column A in Table 2) merely reflects the necessity of using different procedures for the two tables. The two sources of Soviet Interview Project survey data on respondent employment experience are described more fully in the Appendix to the paper.

TABLE 2—PROBIT AND REGRESSION ANALYSIS OF INCIDENCE, FREQUENCY, AND DURATION OF SOVIET UNEMPLOYMENT SPELLS

	A	B	C	D
Dependent Variable:	Experienced Unemployment (0 = No, 1 = Yes)	Number of Spells Reported (1,2,3, etc.)	Multiple Spells (0 = Single Spell, 1 = Two or More)	Duration of Last Spell (Months)
Type of Analysis:	Probit	Linear Regression	Probit	Linear Regression
Independent Variables:				
Sex of Respondent	−.124	−.128	−.463^{a}	.755
Male = 0, *Female* = 1	(.083)	(.193)	(.277)	(.920)
Age of Respondent	−.030^{c}			
End of LNP *in Years*	(.004)			
Age at Last Spell Dummies				
1 = 25 *Years and Above*		.277	.318	−2.50^{b}
		(.224)	(.328)	(1 07)
1 = 35 *Years and Above*		−.140	−.080	.557
		(.269)	(.372)	(1.273)
1 = 50 *Years and Above*		1.263^{c}	1.00^{a}	−.112
		(.432)	(.576)	(2.026)
Educational Attainment:				
1 = *Some Secondary Specialized*	.261^{b}	.214	.114	−.261
	(.133)	(.363)	(.546)	(1.72)
1 = *Any University*	.428^{c}	.092	.146	1.12
	(.122)	(.342)	(.510)	(1.62)
Constant	4.457^{c}	1.184^{c}	4.105^{c}	5.38^{c}
	(.195)	(.344)	(.514)	(1.63)
Cases (N):	2331	113	114	117

Notes: Estimated standard errors in parentheses. Significance levels: ^{a} = 10 percent, ^{b} = 5 percent, ^{c} = 1 percent. The interpretation of the dummy variables for age-groups is the following: coefficients of these dummies reflect the differential effect of being in a specific age-group as compared to the immediately junior age-group. For example, a 46-year-old respondent would be assigned the age-group dummies 1,1,0 for the age-groups 25+ years, 35+ years, and 50+ years, respectively. The interpretation of the dummy variables for educational attainment coefficients of these dummies reflect the differential effect of advanced training of either secondary specialized education or university education compared to having attained a level of education equal to the *attestat* or lower. A complete description of the coding of educational attainment is given in the Appendix.

pay, including military service, for at least six months in a row during a year, that person was coded in the Life History Chart as "employed" during that year. If less than half the year was spent working for pay, then the entire year was coded "not employed." The responses were so coded during the original interview and it is impossible to obtain a finer breakdown. This is insufficient for allocating a respondent's time during the last normal period into different labor force states, for example, all unemployment of less than six months cumulative duration during one calendar year would still have resulted in a classification of "employed" for that year.

To analyze the incidence of a spell of one-month-or-more unemployment for a *labor force participant during the last normal period*, we are forced to classify respondents as either "in the labor force" or "out of the labor force" for the entire last normal period. For the probit sample, we included only respondents with approximately similar labor force participation during the last normal period.[22] Our criterion for including cases for the probit analysis of spell incidence was that at least 70 percent of a respondent's last five years of the last normal period was spent in the labor force (employed or unemployed).[23]

The probit regression on incidence (column A) supports the Table 1 patterns: Younger respondents and college-educated respondents have a higher incidence of unemployment *ceteris paribus*. Holding age and education constant, women may (a t-value of 1.5) have a lower incidence of unemployment. The regressions on multiple spells (column B) fail to yield conclusive *ceteris paribus* patterns except for older respondents (50 years and older) having a higher frequency of multiple spells. Although Table 1 suggests that men and less-highly educated respondents have more multiple spells, these results are not confirmed by multiple regression.

From Table 1 alone, one cannot say to what extent the multiple-spell results are due to the small number of people reporting multiple unemployment spells. For the regression reported in column C, we divided respondents (who experienced unemployment) into those with a single spell (= 0) and those with multiple spells (= 1). As the probit regression shows, older respondents (50 years and older) are more likely and women are less likely to have multiple spells. By dealing with multiple spells in a dichotomous manner, the negative effect of gender (being a woman) on multiple spells is strengthened.

The regressions on duration of unemployment spells (of those reporting unemployment) show that young respondents (under 25 years of age) have longer spells of unemployment than workers 25 and over. Although Table 1 suggests women have a longer duration of unemployment, regression D in Table 2 does not support that conclusion.

The regressions of Table 2 reveal the reasons for younger workers having higher unemployment rates. They are more likely to have at least one spell of unemployment, and the duration of the spell is longer *ceteris paribus*. College-educated workers have higher unemployment rates simply because they are more likely to have experienced an unemployment spell. There are no significant differences in multiple spells and duration of spells according to education levels.

It is possible to use the regression coefficients reported in Table 2 (columns A, B, and D) to obtain an exact algebraic breakdown of the effects of gender, age, and education on the unemployment rate through their effects on unemployment incidence, number of spells, and spell duration. Predicted values of incidence, number of spells and duration for 24 groups cross-classified by sex, age-group, and educational attainment (2 by 4 by 3 classification) were calculated.[24] Next synthetic unemployment rates for the 24 groups were then calculated using equation (1). Taking logs of both sides of equation (1) we have

$$(2) \qquad \ln(U) = \ln(t) + \ln(N) + \ln(AD) + \ln(1/60).$$

---

[22] For example, consider a young person who completed education during the first three of the five years in the last normal period, followed by only two years of labor force activity. If during those remaining two years a spell of unemployment occurred, the incidence of unemployment is really much higher than it is for someone who experienced unemployment but who spent the entire five years of the last normal period in the labor force (1 spell/two years > 1 spell/5 years).

[23] Specifically, our sample for the probit regression includes respondents for whom the sum of the number of "employed" years out of the last five years (in the sense of the Life History Chart) *and* months in a long (six-months-or-more) unemployment spell exceeded three and a half years.

[24] For the probit regression of incidence we used 22, 30, 42, and 55 years for the < 25 years, 25–34, 35–49, and 50 ≤ age-groups, respectively. Predicted incidence was calculated as $\Phi(X\hat{\beta} - 5)$, where X is the matrix of independent variables, $\hat{\beta}$ are the coefficient estimates from column A of Table 2, and $\Phi(\ )$ is the cumulative standard normal distribution function.

TABLE 3—EFFECTS OF GENDER, AGE, AND EDUCATION ON UNEMPLOYMENT RATES, INCIDENCE, NUMBER OF SPELLS AND SPELL DURATION

	Dependent Variable: Natural logarithm of			
	Unemployment Rate	Incidence	Number of Spells	Duration
Female	−.142	−.240	−.084	.182
25+ Years	−.738	−.397	.206	−.547
35+ Years	−.633	−.685	−.099	.151
50+ Years	−.237	−.868	.660	−.029
Some Secondary Specialized Education	.598	.528	.139	−.069
Some University	1.149	.836	.062	.251
Intercept	−4.526	−2.244	.171	1.641

Note: The coefficient for an independent variable listed in the unemployment rate regression is equal to the sum of the last three coefficients in the same row. The intercept in the unemployment rate regression also includes $\ln(1/60) = -4.094$. For interpretation of dummy variables see the notes to Table 2.

Each of the transformed variables of (2) were separately regressed on the same matrix of sex, age-group, and educational attainment dummies for the 24 groups. As can be readily verified by examining the rows of Table 3, the sum of the coefficients for a particular dummy variable in the incidence, number of spells, and spell duration regressions is exactly equal to the corresponding coefficient in the synthetic unemployment rate regression.[25]

The coefficients presented in Table 3 supplement both Tables 1 and 2. The first column of Table 3 provides information on the relative impact of gender, age, and education (this time *ceteris paribus*) on the Soviet unemployment rate. Comparing the coefficients in the first column of Table 3, we confirm that the impact of gender on the Soviet unemployment rate is relatively minor. We are also able to confirm the relatively strong (negative) effect of age and (positive) effect of increased educational attainment found in the one-way breakdowns of the unemployment rate in Table 1. Comparing the signs of the coefficients in the last three columns in Table 3 with the breakdowns of Table 1, one finds overwhelming agreement. The only sign reversal is for the minor drop in the number of spells associated with higher educational attainment in Table 1. This is not particularly important since the estimated coefficient of increased educational attainment in the regression on the number of spells was statistically insignificant.

Table 3 is also interesting because the effects of each factor on incidence, number of spells, and spell duration have been rendered comparable. For example, moving from the 25 years-and-older category to the 35 years-and-older category reduces the unemployment rate (a coefficient of −.633). This negative coefficient is "explained" by the lower incidence associated with the older age category (−.685), offset somewhat by a slightly longer spell duration (.151). The substantially higher unemployment associated with some university education is "explained" primarily by the higher incidence of unemployment of those having received some university education, augmented by a somewhat longer duration of completed spells.

---

[25] This decomposition is similar to that used by Orley Ashenfelter and John Ham, 1979, to show the effect of schooling on earnings, unemployment, and wage rates for a sample of U.S. white males. Ashenfelter and Ham were able to run their regressions on *individual* observations. It is necessary for us to run regressions on grouped data because 1) our data are inadequate to calculate unemployment rates for the individual respondents, and 2) grouping the data keeps the variables of formula (1) strictly positive (i.e., no zeros), which permits the log transformation in equation (2).

The major conclusion to be drawn from Table 3 is that unemployment rate differences across subgroups of the sample are accounted for primarily by differences in incidence rather than by differences in the number or duration of spells.

V. Comparisons with U.S. Incidence and Duration Data

Comparisons of Soviet unemployment patterns with those of industrialized capitalist countries shed light on unemployment under radically different institutional settings. Table 4 provides comparable Soviet (1974–1979) and adjusted U.S. (1967–1977) data on *one-month-or-more* unemployment incidence, duration of completed spells, and unemployment rates by age and sex. We present data for white Americans because of the homogeneity of the Soviet Interview Project sample (urban residents from European Russia, 90 percent Jewish). We are only able to provide directional information for education. Although we have not been able to make comparable adjustments to unemployment statistics broken down by demographic groupings for other major capitalist countries, available evidence indicates that U.S. patterns are shared by other countries.[26]

A. *Unemployment and Age*

Mature workers in both the Soviet and U.S. labor markets have lower unemployment rates than young workers. In both cases this is accounted for by younger workers having a higher incidence of unemployment spells. While for American workers there is a gradual upward trend in the duration of completed spells with age, we do not find such a positive association between the duration of completed spells and age for the Soviet Interview Project sample—indeed, we find a significantly longer average duration for workers under the age of 25.

Given the elaborate Soviet administrative machinery for placing young people in jobs,[27] our finding of a relatively higher unemployment incidence (and longer duration) for younger Soviet workers was not entirely anticipated. The greater use of official placement for young workers in the Soviet Union is apparent in Table 5. For young workers it is the second most important source of finding jobs, but it is of little importance for mature workers. While young and mature U.S. workers use basically the same job-search methods, young Soviet workers find their jobs through official "matchmaking" and mature Soviet workers primarily use private matchmaking (networking). Furthermore, "job search" seems better to characterize U.S. workers (who to a very large extent find jobs by directly contacting employers). The main point here is that Soviet official placement has not completely offset the strong tendency of young people to have a higher incidence of unemployment (due to less experience at finding jobs and "shopping for jobs" associated with career search).

B. *Unemployment and Gender*

In the Soviet sample, men and women have very similar unemployment rates. While the Table 2 coefficients indicate higher incidences and shorter spell duration for men, neither result is particularly significant (t-statistics of -1.5 and .82, respectively).[28] This basic similarity in the unemployment of Soviet men and women contrasts with U.S. experience. U.S. women have higher one-month-or-more unemployment rates than do U.S. men (Table 4). The decomposition of the unemployment rates into incidence and duration components shows that U.S. women have a higher incidence of spells than men, but about the same average duration.

---

[26] Compare Wolfgang Franz and Heinz König, 1986, for West Germany as well as the other papers from a 1985 conference on unemployment edited by Charlie Bean, Richard Layard, and Stephen Nickell, 1986. A valuable collection of statistics on the distribution of unemployment by duration of interrupted spells and demographic groups can be found in OECD, 1986, pp. 74–77.

[27] On this, see Nick Lampert, 1986.

[28] U.S. Department of Labor, 1977, p. 24, and Table 3 show that men have a greater frequency of *multiple* spells than do women in both the United States and the Soviet Union.

TABLE 4—PATTERNS OF UNEMPLOYMENT BY AGE AND SEX, ONE-MONTH-OR-MORE SPELLS FOR THE UNITED STATES (WHITE ADULTS, 1967–77) AND THE USSR (SOVIET INTERVIEW PROJECT, 1974–79)

	Incidence (Monthly Spells per 100 Members of Labor Force)		Average Duration (Months)		Unemployment Rate (Percent)	
	USSR	United States	USSR	United States	USSR	United States
Total	.23	1.3	4.8	3.1	1.1	3.9
Men	.28	1.1	4.4	3.1	1.2	3.5
Women	.19	1.5	5.4	3.0	1.0	4.5–4.6
18–24 Years	.59		6.0		3.5	
25–34 Years	.35		4.1		1.5	
35–49 Years	.11		4.3		.5	
50+ Years	.14		4.0		.5	
16–19 Years		3.4–3.6		2.6		8.9–9.4
20–24 Years		2.1		3.1		6.4
25–59 Years		.9		3.2		2.9
60+ Years		.7		3.9		2.8

Source: USSR, calculations from Table 1 with incidence figures converted to an annual basis. U.S. figures estimated by authors using transition probabilities for demographic groups estimated from CPS gross flows data by Ronald Ehrenberg, 1980, and official CPS data on employment and unemployment. The adjustment procedure to exclude completed spells of less than one-month duration for U.S. unemployment statistics is described in the Technical Appendix of Gregory and Collier, 1987, pp. A17–A27.

TABLE 5—JOB SEARCH METHODS USED BY YOUNG AND MATURE WORKERS OF THE SOVIET INTERVIEW PROJECT (SIP) SAMPLE AND UNEMPLOYED U.S. WORKERS (PERCENT)

	Young Workers		Mature Workers	
	USSR (SIP)	United States	USSR (SIP)	United States
Respondent Inquired at Enterprise/Institution	11	49	15	41
Told About Vacancy by Friend/Relative/Teacher	52	9	63	10
Public Announcement or Advertised at School	10	19	14	19
Official Placement[a]	26	19	7	25
Other	1	3	1	5
	100	100	100	100

Source: Soviet Interview Project General Survey; U.S. Department of Labor, 1983, p. 84. For the United States "young workers" are 19–24 years of age and mature workers are those aged 35–54.

[a] For SIP "official placement" includes the responses "did practicum/special training there," "komsomol placement," "was invited/recruited," and "military commission."

To explain the similar rates of male and female unemployment in the Soviet Union, one needs to explain the relative (i.e., compared to U.S. women) lack of movement between labor force states by Soviet women. U.S. women move in and out of the labor force with greater frequency than men. In contrast Soviet women appear to have less, certainly no greater, movement between labor market states than do Soviet men.[29]

Soviet women have been subject to antiparasitism laws as have men. Even mothers with young children have the obligation to work outside the home. This has fostered an attitude that complete specialization in household production is simply not a legitimate option. Moreover, Soviet women must carry out household production in a world of disequilibrium prices (including substantial queuing time). Soviet women deal with disequilibrium by "stealing" relatively more time from their jobs than men.[30] Stability in household production therefore depends on informal understandings worked out between the female employee and her immediate superior for taking time off from work for personal business. Each job change would require a Soviet woman to establish a new set of "implicit contracts."

C. *Unemployment and Education*

The most striking contrast between Soviet and U.S. unemployment patterns is the positive relationship between education and unemployment in the Soviet Union and the negative relationship found in the United States. We do not have comparable one-month-or-more unemployment data by level of education for the United States, but the available evidence indicates that highly educated American workers have a lower incidence of unemployment, fewer multiple spells, and about the same duration of completed unemployment spells.[31] In contrast, highly educated Soviet workers have a higher incidence of unemployment and (possibly) longer duration spells.

The inverse relationship between education and unemployment in the United States has not been dealt with satisfactorily in the literature. It is not our task to explain the inverse U.S. relationship. More important is an explanation for the observed positive relationship between education and unemployment in the Soviet Union.

Several factors can be advanced to explain the positive relationship between education and unemployment in the Soviet Union. First, higher-education graduates in particular are administratively assigned their first jobs. Presumably this official placement system is less likely to produce a successful match between employee and employer than one allowing greater choice by job candidates. University graduates are often assigned to remote regions, whereas public school graduates usually expect to stay in their hometowns. Upon fulfillment of the formal obligations of the first placement, the university graduate's search for the next job will require turning to less formal networks of job information.[32] As these young workers hunt for new jobs which more closely match their preferences (especially geographic preferences), they are forced to turn to unofficial channels. Second, there is evidence that Soviet authorities have overinvested in human capital, thereby creating a mismatch between the number of college graduates and the number of jobs available.[33]

---

[29] See the paper by Elizabeth Clayton, 1986.

[30] See Gregory, 1987.

[31] These conclusions come from the Work Experience Survey series which show incidence, multiple spells, and duration by detailed occupations. Occupations that require higher levels of education consistently show lower incidence and fewer multiple spells. The CPS unemployment series also points to an inverse relationship between unemployment and education in the United States.

[32] For an American journalist's account of the placement process (*raspredeleniie*) and its significance for the individual graduate awaiting job placement, see David Shipler, 1984, pp. 163–79.

[33] Granick, 1987a, b presents evidence of overinvestment in human capital. For a survey of the allocation mechanisms for labor in the Soviet Union, see Gregory and Robert Stuart, 1986, pp. 226–38.

VI. Unemployment Rate of the Referent Soviet Population

We turn from demographic patterns of unemployment to international comparisons of national unemployment rates. To detect patterns of unemployment, it was not necessary for the Soviet Interview Project sample to have distributions of demographic characteristics identical to the referent Soviet population. However, care must be taken when inferring the unemployment rate of the referent population. Only if the Soviet Interview Project sample and the referent population had identical distributions with respect to the characteristics relevant to unemployment would the rate be representative of the referent Soviet population (allowing for sampling error). The age and gender distributions of the Soviet Interview Project sample are virtually identical to the referent population, but the sample is more highly educated than the referent population.[34] Therefore one must calculate a hypothetical unemployment rate imposing the educational distribution of the referent population on the Soviet Interview Project sample. Such a calculation (which includes an adjustment for memory decay) gives a 1.2 percent unemployment rate for the referent Soviet population which we have entered in Table 6.[35] This exercise shows that the effect of reweighting is small and confirms that recall biases are more serious than sample bias. The residual uncertainty of our estimate is whether being Jewish affects the probability of unemployment independently of other factors for which we are able to control. Respondents may have been more sensitive to antiparasitism laws and have greater concern for a record of uninterrupted employment. However, being Jewish may have made them more vulnerable to arbitrary firings as well. Without a truly random sample from Soviet society, there is no way to determine the net effect of these two distortions. In our view, it is unlikely that a "pure Jewish distortion" (if present at all) would work to substantially alter the Soviet unemployment rates calculated in this paper.[36]

VII. Previous Estimates of Soviet Unemployment

Previous attempts to estimate Soviet unemployment rates, Peter Wiles, 1972, and Granick, 1987a, were based on partial data on turnover rates and duration between jobs from published official Soviet sources. Table 6 presents the estimates of Granick and Wiles which measure the *total* Soviet unemployment rate, whereas the Soviet Interview Project estimate measures only one-month-or-more unemployment. Wiles calculated a 1962–63 total Soviet unemployment rate of 1.3 percent (with a range of 1.0–1.8 percent) from isolated scattered turnover data. Granick, working with more complete Soviet data on turnover and separations, calculated

---

[34] The Soviet Interview Project (SIP) and referent population age, gender, and education distributions are as follows:

		SIP	Referent Soviet Population
A. By Age, percent			
(SIP age upon arrival)	21–30 years	21.6	24.7
	31–40	25.6	25.9
	41–50	21.2	21.5
	51–60	15.7	15.4
	61 and over	15.9	12.4
		100	100
B. By Sex, percent			
	Men	43.4	44.3
	Women	56.6	55.7
		100	100
C. By Educational attainment, percent			
Any higher education		41.9	27.3
Some secondary specialized or *attestat*		40.7	40.6
Lower than *attestat*		17.4	31.1
		100	100

[35] We reaggregated the unemployment rates for each of the educational categories from Table 1 using first the Soviet Interview Project sample educational distribution and then the referent population's educational distribution above. The ratio of the two reaggregated unemployment rates (.93) is applied to the Soviet Interview Project unemployment rate (1.1 percent) of Table 1 to yield a hypothetical unemployment rate (1.02) for the referent population, which is adjusted upward using the memory-decay factor to 1.2 percent.

[36] Ofer and Vinokur, 1984, discuss the likelihood of such a distortion affecting wages at greater length.

TABLE 6—COMPARISONS OF UNEMPLOYMENT RATES: USSR AND SELECT CAPITALIST ECONOMIES (PERCENT)

	Year(s)	One-Month-or-Longer Spells	All Spells
Soviet Union			
Soviet Interview Project[a]	1974–79	1.2	
Wiles (1972)	1962–63		1.1–1.8
Granick (1987a)	late 1970s		1.5–3.0
Capitalist Economies			
Low Unemployment			
United States			
a. All Adults	1969	1.9	2.7
b. White Adults	1969	1.7	2.4
West Germany	1971–72	.5	1.0
Japan	1980	1.9	2.0
High Unemployment			
United States			
a. All Adults	1976	5.6	6.5
b. White Adults	1976	5.1	5.9
West Germany	1984	8.0	8.2
United Kingdom	1980	6.1	6.8

[a] Population of medium-to-large cities, European USSR.

Note: Western unemployment statistics have been adjusted using empirical survivor functions. For the United States, U.K., and Japan the survivor functions were estimated from data on the distribution of the duration of interrupted spells in U.S. Department of Labor, 1982, 1983, and OECD, 1986. For Germany the survivor function was estimated using the average duration of completed spells from Franz and König, 1986. The adjustments needed to sweep out spells of less than one-month duration are described fully in the Technical Appendix of Gregory and Collier, 1987, pp. A17–A31.

a range for the Soviet unemployment rate in the late 1970s of 1.5 to 3.0 percent with a point estimate of 2.3 percent. One must resist the temptation to draw inferences about trends in the Soviet unemployment rate from the Wiles and Granick estimates (which appear to show an upward trend). The two estimates use different methodologies, and Granick's statistics from the 1970s are much richer than Wiles data for the early 1960s. Unlike the U.S. case, where there has been a clear secular increase in the natural rate of unemployment, there are no strong demographic reasons to expect such an upward drift in the Soviet unemployment rates.[37]

We repeat that we regard the Soviet Interview Project-based unemployment figure of 1.2 percent as a lower-bound estimate of one-month-or-more unemployment. We cannot offer an upper bond, but the lower bound presented here sheds some light on the Wiles and Granick estimates. First, our estimate for 1975–79, which omits unemployment spells of less than one month, is roughly the same as the Wiles estimate of the *total* Soviet unemployment rate for the early 1960s. We cannot know by what factor our estimate must be raised to yield an estimate of the

---

[37] The proportion of women in the labor force has remained stable, and the aging of the Soviet labor force has meant a decline in the employment share of young workers. The major shift has been the rising proportion of university graduates in the labor force, which rose from 6 percent in 1960 to 11 percent in 1980. As noted in this paper, the rising share of university graduates could have meant a very slight upward trend in Soviet unemployment, but the falling share of young workers would serve to offset this effect.

total Soviet unemployment rate, but even without adjustment it falls within Wiles confidence interval. Our rate falls within Granick's suggested confidence interval with only a minor upward adjustment for short-term unemployment. This consistency with the two previous estimates derived from Soviet sources is reassuring.

VIII. International Comparisons

In the bottom panel of Table 6 we have assembled unemployment rates for the United States, West Germany, Japan, and the United Kingdom for comparisons with the three estimates of Soviet unemployment discussed above. The four countries reasonably cover the spectrum of unemployment experienced in major industrialized capitalist countries. For the United States and West Germany we have sufficient data to calculate estimates of one-month-or-more unemployment rates during a peak and a trough of the business cycle. For Japan (low unemployment) and the United Kingdom (high unemployment), we are able to calculate one-month-or-more unemployment rates only for 1980. For each country, the corresponding official unemployment rate (which includes spells of all durations) is given in the last column of the table.

The one-month-or-more unemployment rates presented in Table 6 are meaningful in their own right. An unemployment rate that sweeps out the shortest spells of unemployment will capture much of what Akerlof and Yellen, 1985, consider salient unemployment —unemployment that is likely to be remembered by those experiencing it and unemployment that involves more than a relatively uneventful, brief transition between jobs.[38]

Table 6 reveals that the unemployment rate we have estimated for the Soviet referent population is relatively low, especially when compared with the high unemployment figures for the United States, West Germany, and the U.K. at the bottom of the table. Indeed it is clear from looking at the low-unemployment and high-unemployment rates for the capitalist economies listed in the table that an "average capitalist" rate would be several times larger than the 1974–79 average for the Soviet urban population. While this average difference comes as no surprise, it is quite interesting to see that the unemployment rates experienced by capitalist economies during good times are quite similar to the estimated Soviet rate. Indeed, the West German rate for 1971–72 is even lower than the Soviet estimate.

IX. Conclusions

Data from a sample of former Soviet citizens provide a first opportunity to estimate Soviet unemployment using standard survey methodology. While Western researchers (and even Soviet scholars) have focused on frictional unemployment, the Soviet Interview Project sample allows us to analyze longer-duration unemployment. Our research places a lower bound on the Soviet unemployment rate (excluding spells of less than one month) of 1.2 percent. This Soviet rate for the mid-1970s is well below comparable one-month-or-more unemployment rates we have estimated for selected industrial capitalist countries for the same period. Soviet long-duration unemployment rates, however, are not unique. In good years West Germany and Japan have achieved similar rates.

This study helps to reveal the effect Soviet institutions have on patterns of unemployment. The most striking difference between U.S. and Soviet patterns of unemployment is the positive relationship between education and unemployment in the Soviet Union. Although we cannot offer a conclusive explanation for this unusual Soviet pattern, we believe both the official placement system of college graduates and Soviet manpower planning contribute to the reversal. A second striking difference is that Soviet women move between labor market states less frequently than U.S. women. We are unable to find significant differences between Soviet male-female unemployment patterns, whereas gender differences are quite apparent in the U.S. case. Finally, our estimates indicate that Soviet labor market institutions have not completely offset a common tendency for unemployment rates to be higher for younger workers. Despite an extensive system of administrative placement for new workers (which we have shown is used primarily by the young), there still exists an inverse relationship between age and unemployment in the Soviet Union.

---

[38] For U.S. data we have examined the relationship between Akerlof and Yellen's ratio of "salient" unemployment (from the WES) to total unemployment (from the CPS) and our estimates (for example, in Table 6) of the ratio of one-month-or-more unemployment to total unemployment. The two measures were found to move closely together over the business cycle (both ratios increase with downturns in economic activity).

If we wish to attribute the unusual unemployment patterns of this paper to the Soviet system, it would be interesting to discover whether the distinctly Soviet patterns disappear once the Soviet emigrants have entered the U.S. labor market. U.S. Census data for 1980[39] show that Soviet emigrant women who had entered this country between 1975 and 1980 had distinctly higher unemployment rates than Soviet emigrant men. We find that unemployment rates fall with age up to the age-group 35 to 39 years, but rise thereafter. However, we are unable to conclude from the 1980 census data that the change in patterns truly demonstrates the impact of economic systems on labor force behavior. The census data capture the Soviet emigration cohort analyzed in this paper still adjusting to the new economic system. Nonetheless, the differences are suggestive and point to a promising avenue for future research.

APPENDIX

In this appendix we provide a general description of our methods to calculate unemployment rates from the Soviet Interview Project General Survey. Full documentation can be found in Gregory and Collier, 1987, which also documents the calculation of the one-month-or-more unemployment rates for the Western economies presented in Tables 4 and 6.

A. *The Formula for Calculating an Unemployment Rate from the Soviet Interview Project General Survey Responses*

The Soviet unemployment rates in this paper are average rates for the final five years of the last normal period of the respondents' lives in the Soviet Union. All variables listed below refer to the sixty months of the last normal period for each respondent.

Notation:

D_{ij} = Number of j-month long spells of the ith respondent;

Z = Number of respondents who reported experiencing an unemployment spell of 30 days or more;

N = Number of respondents (2793);

i = Index for respondents;

j = Index for duration of spell, that is, $j = 2$ refers to a completed spell of two months;

E_i = Months employment of the ith respondent;

W_i = Dummy variable = 1 if respondent working at end of last normal period;

u_{it} = Dummy variable = 1 if ith respondent unemployed in month t;

e_{it} = Dummy variable = 1 if ith respondent employed in month t;

l_{it} = Dummy variable = 1 if ith respondent in labor force during month t.

Derivation of formula for unemployment rate:

Using the above notation, we have the following:

(A1) $\sum_j j \cdot D_{ij}$ = Total months of unemployment of ith respondent;

(A2) $\sum_j D_{ij}$ = Number of spells of the ith respondent;

(A3) $\sum_i \sum_j j \cdot D_{ij}$ = Total months of unemployment for the sample;

---

[39] Unemployment rates (in the United States) of Soviet emigrants for 1980 were calculated from cross tabulations of labor force status of immigrants by country of origin, date of arrival, sex, and age from 1980 Census microfiche, U.S. Department of Commerce, 1985.

Soviet Emigrants' 1975–80 Arrival	Unemployment Rates in the United States, 1980, percent
Total (20–64 years)	12.9
Men	10.8
Women	15.6
20–24 Years	13.0
25–29	11.6
30–34	11.6
35–39	9.6
40–44	14.8
45–49	13.1
50–54	15.1
55–64	16.4

$$\text{(A4)}\qquad \sum_i \sum_j D_{ij} = \text{Total number of spells of the sample.}$$

Divide (A3) by (A4) to obtain the average length of a spell of unemployment for the sample:

$$\text{(A5)}\qquad \frac{\sum_i \sum_j j \cdot D_{ij}}{\sum_i \sum_j D_{ij}} = \text{Average length of a spell;}$$

$$\text{(A6)}\qquad \frac{\sum_i \sum_j D_{ij}}{Z} = \text{Average number of spells of those with unemployment;}$$

$$\text{(A7)}\qquad Z/N = \text{Proportion of sample reporting unemployment.}$$

We determine the average months of unemployment per respondent by multiplying (A7) by (A6) by (A5):

$$\text{(A8)}\qquad \frac{\sum_i \sum_j j \cdot D_{ij}}{N} = \frac{Z}{N} \cdot \frac{\sum_i \sum_j D_{ij}}{Z} \cdot \frac{\sum_i \sum_j j \cdot D_{ij}}{\sum_i \sum_j D_{ij}}.$$

The last term in the product (average spell duration) cannot be calculated directly since respondents were only asked about the duration of their *last* spells of unemployment. We have simply used the average duration of the last spells of unemployment as an estimate of average spell duration (only one-third of respondents with unemployment had multiple spells).

The average number of months of labor force activity (defined as time employed or unemployed) remains to be determined.

$$\text{(A9)}\qquad E_i + \sum_j j \cdot D_{ij} = \text{Months } i\text{th respondent in labor force.}$$

Summing over i respondents and dividing by N we obtain the average time in the labor force

$$\text{(A10)}\qquad \frac{\sum_i \left(E_i + \sum_j j \cdot D_{ij} \right)}{N} = \text{Average number of months each respondent is in the labor force.}$$

The one-month-or-more unemployment rate is simply (A8) divided by (A10):

$$\text{(A11)}\qquad U = \frac{\sum_i \sum_j j \cdot D_{ij}}{N} \Bigg/ \frac{\sum_i \left(E_i + \sum_j j \cdot D_{ij} \right)}{N} = \frac{\text{months unemployment}}{\text{months in labor force}},$$

where the numerator is calculated as in (A8).[40] E_i is not known for each respondent.[41] In place of the average number of months of employment in the denominator, we use the employment rate of the sample at the end of the last normal period times 60 months, that is, we assume a constant employment rate over the last five years for each subgroup given in Table 1.

$$\text{(A12)}\qquad \left(\sum_i E_i / N \right) \text{ is estimated by } 60 \cdot \left(\sum_i W_i / N \right),$$

where W_i is a dummy variable equal to one if the ith respondent was working at the end of the last normal period.[42]

B. *Estimating Unemployment Rates for Soviet Interview Project Sample*

178 of the 2793 respondents answered yes to the question (*UNEMP5YR*),[43] whether they had experienced at least a 30-day spell of unemployment during the last five years of the last normal period. Since the averages of duration and number of spells used in formula (A11) are based solely on the respondents with unemployment experience, it was feasible to examine individual cases both for problem cases and for assigning responses to appropriate age categories. Over 100 respondents required special handling or adjustments of some kind. In this section we describe these adjustments.

Duration of unemployment and number of spells:

1. The measure of duration (*DURUNE*) was calculated from the coded responses for the beginning and end months and years of the last spell of unemployment. For spells with *DURUNE* greater than or equal to two months, this is equal to the expected duration of the spell (assuming independence of the months of beginning and ending unemployment). However, this measure is a biased estimate for *DURUNE* equal to

---

[40] In the Technical Appendix of Gregory and Collier, 1987, we show that equation (A11) in our Appendix, is a weighted average of monthly one-month-or-more unemployment rates.

[41] This limitation of the data needs to be emphasized. Had we known the labor force status of all respondents during each of the sixty months of the last normal period, unemployment rates could have been calculated directly and this appendix would have been much simpler and shorter.

[42] To be excruciatingly correct we should include the second term in the denominator of equation (A11), $(\sum_j j \cdot D_{ij}) \div N$. We ignored this adjustment in our calculations because of the relatively small magnitude of this number (about 2 weeks) compared to the error from estimating the months of employment (a magnitude of about 40 months).

[43] Wherever possible we give the exact variable name from the SIP tape in uppercase characters.

zero or one. For example, if the spell began and ended in the same month and was at least 30 days long, then the unbiased estimate of the duration of the spell would be 1 month (not zero as calculated).[44] Similarly for a spell which begins and ends in adjacent months, conditional on the spell being at least 30 days long, the expected length of the spell is 4/3 and not one month. We recoded the thirty cases with *DURUNE* coded as 1 month with 1.33 months.

2. We have corrected *DURUNE* for seventeen spells which fell completely outside the reference period and for spells which crossed the end points of the last normal period.[45]

3. There are twenty-seven respondents with incomplete or incorrectly coded responses for duration of last spell. Thirteen of those respondents have either zero or twelve months coded for *DURUNE*, but the first and last months are suspiciously coded June (see 44 above). These missing data are another reason for using our formula (A11), which is calculated using averages rather than individual responses.

4. There are ten respondents whose recorded end of last spell of unemployment exactly coincides with the final month of the last normal period. For four respondents who reported less than a year of unemployment, we treated those reported months similarly to the months in the sixty months of the last normal period of respondents whose spells crossed the end of the last normal period. The other six respondents in all likelihood confused the state of unemployment with being out of the labor force and were omitted from calculations of average duration (which would have been significantly affected by their inclusion).

5. For the calculation of average duration by last spell, we were able to assign the (last) spell to the exact age of the respondent at the time of the spell. For the calculation of average number of spells within a five-year period it was necessary to allocate spells and sample years to different age-groups for those respondents whose five-year reference periods fall in two age-groups.[46] 53 respondents had their multiple spells and/or time allocated between age-groups in this manner. Years of experience and number of spells were summed for each age, sex, and educational attainment group.

Labor force participation and incidence of unemployment:

There are two sources of labor force participation data in the Soviet Interview Project survey. 1) The variable *DOINGLNP* gives the respondent's answer to the following question: "In (the end of the last normal period), what were you doing most of the time—working, going to school, keeping house, or something else?" 2) The Life History Chart[47] also yields limited information on the respondent's employment for all five years of the reference period (after the age of 18).

For the calculation of average rates in Table 1, we used the variable *DOINGLNP* because we believe it to be more accurate than the Life History Chart which misses periods out of the labor force (with or without a spell of unemployment) shorter than six months. For the probit regressions of unemployment incidence in Table 2, we were forced to use data from the Life History Chart since we are dealing with individual observations and not with aggregated data as in Table 1.

Incidence by age-group:[48]

From a simple cross tabulation of *UNEMP5YR* and a respondent's age at the end of the last normal period, we have something close to, but not exactly, incidence by age. Because the observation window of five years is so wide, it was necessary to adjust the data to correspond to individual years of age. A 25-year-old respondent is reporting on own unemployment experience for ages 21 through 25, a 26-year-old for ages 22 through 27, etc.

Step 1. We cross tabulated labor force participation rates using *DOINGLNP* (end of last normal period) and incidence of unemployment (entire five-year period) using adjusted *UNEMP5YR* with respondent's age at the end of the last normal period.

---

[44] We found that all the "zero" duration one-month-or-more spells (eight cases) were the consequence of recoding "unknown" and "missing" values of month values to the month of June.

[45] Eleven spells were completed after the final month of the last normal period. That portion of unemployment which occurred after the last month reflects some of the change in the respondent's life and is not believed to be representative. Hence these spells were not included for calculating average duration but were included for calculating incidence of unemployment. Six spells began before the reference period and ended during the 60 months of the last normal period. We included these earlier spells in calculating both average duration and incidence.

[46] For example, a respondent who was 28 years of age at the end of the last normal period contributed two years to the 18 to 24-year group and 3 years to the experience total of the 25 to 34-year group. The reported 3 spells of unemployment for this individual were allocated: 1 spell to the 18 to 24-year group and 2 spells to the 25 to 34-year-old group.

[47] The Soviet Interview Project Life History Chart was constructed from the responses to the following question: "Now think about the times in your life after you reached the age of 18 when you were not working for pay for at least six months in a row. You may have been in school, (having or) taking care of children, or doing something else. Look at this chart again, and tell me in which years you were not working for pay for at least six months. (Time in the Armed Forces counts as 'working for pay'.)"
The [T–18] responses (0 = if less than 6 months in a row working for pay, 1 = if 6 or more months in a row spent working for pay) constitute a crude employment history for a respondent who was age T at the end of the last normal period.

[48] A numerical illustration of the procedure outlined in this section is given in Gregory and Collier, 1987, pp. A13–A15.

Step 2. Average annual incidence per respondent over last 5 years was calculated as follows: Let p_1 and p_5 be the one-year probability of experiencing a spell of unemployment and the probability of experiencing a spell of unemployment during a five-year period, respectively. p_5 is taken from step 1. The probability of not experiencing unemployment five years in a row (assuming independence of the probabilities across years) is $[1-p_5]=[1-p_1]^5$, which we solve for $p_1 = 1-(1-p_5)^{1/5}$.

Step 3. For each birth cohort of respondents the average annual incidence (p_1 from step 2) was treated as an "observation" for the ages $a_0, a_{-1}, a_{-2}, a_{-3}, a_{-4}$ (where age at the end of the last normal period is defined a_0, next to last is a_{-1}, etc.). For each age (a), we calculated a weighted average of the average annual probabilities for the five cohorts which experienced that age during the five years of the last normal period. The weights are the number of respondents in each cohort of respondents.

Step 4. In order to estimate the incidence of unemployment per worker by age, we generated the expected number of people experiencing unemployment by age and divided that number by the expected number of people working by age. We multiplied the number of respondents who passed through a particular age during the last five years of the last normal period by the age-specific participation rate (step 1) and age-specific unemployment incidence (step 3).

Step 5. Annual incidence of unemployment per employed person by age-groups was obtained by simply adding the expected number of unemployed for the individual ages for each of the aggregate age-groups and dividing that sum by the sum of the expected number of people working by the individual ages—both from step 4. The annual probabilities (p_1) for the age-groups thus obtained are transformed into five-year incidences p_5 using the formula $1-(1-p_1)^5 = p_5$. (Compare step 2.)

Incidence by sex:

For the estimated incidence of unemployment by sex, steps 1 through 4 described above were followed. Because of the relative infrequency of unemployment in the Soviet Interview Project sample, the age-specific average rates over the last five years of the last normal period calculated in step 2 showed a large variation for each of the sexes. The weighted average calculated in step 3 smoothed the rates across ages. Step 5 was replaced by aggregating expected workers and expected people who experienced unemployment by age for men and women separately.

Incidence by educational attainment:

A variable for the respondent's highest level of education (*REDHIC*) was constructed by Soviet Interview Project researcher Michael Swafford. We have reaggregated *REDHIC* into three categories.[49]

Even the three-way classification of education leads to many empty cells in a cross tabulation of unemployment experienced by respondent's age and by the respondent's highest educational attainment. Before step 1 we smoothed the frequency counts in the separate cross tabulations of all respondents, those working at the end of the last normal period, and adjusted *UNEMP5YR* by age and educational attainment. The frequency counts were smoothed using a three-age (in years) moving average.[50] Then as in our estimate of unemployment incidence by sex we followed steps 3 and 4 above. In step 5 we aggregated by the three educational categories.

---

[49] *Attestat or less*: (*REDHIC* 10 or less) "eight years of general schooling or less"; "at least one year of trade (vocational-technical) school"; "more than 8 years of general schooling or at least one year of secondary specialized school"; "completed at least a two-year trade-school program or attended trade school at least three years"; or "earned an *attestat zrelosti* or equivalent by any means."

Some secondary technical: (*REDHIC* 11 or 12) "earned an *attestat* and attended at least two years of trade school" or "earned an *attestat* and attended specialized secondary school at least one year."

Some higher education: (*REDHIC* 13 through 19) "attended an institution of higher education for at least one year"; "completed a higher education"; "engaged in some graduate study (*aspirantura*)"; or "earned an advanced degree."

[50] For a numerical illustration see Gregory and Collier, 1987, p. A17.

REFERENCES

Akerlof, George A. and Yellen, Janet L., "Unemployment Through the Filter of Memory," *Quarterly Journal of Economics*, August 1985, *100*, 747–73.

Anderson, Barbara and Silver, Brian, "Descriptive Statistics for the Sampling Frame Population," *Soviet Interview Project Working Paper*, No. 2, 1984.

——— **and** ———, (1987a) "The SIP General Survey Sample," in James Millar, ed., *Politics, Work, and Daily Life in the USSR*, New York: Cambridge University Press, 1987, 354–71.

——— **and** ———, (1987b) "The Validity of Survey Responses from Interviews of Married Couples in a Survey of Soviet Emigrants," *Social Forces*, December 1987, *66*, 537–54.

———, ——— **and Lewis, Robert,** "Demographic Estimates for the Post-Sampling

Weights of the SIP General Survey, *Soviet Interview Project Working Paper*, No. 4, revised 1986.

Ashenfelter, Orley and Ham, John, "Education, Unemployment, and Earnings," *Journal of Political Economy*, October 1979, *87*, S99–S116.

Bean, Charlie, Layard, Richard and Nickell, Stephen, eds., *Unemployment Supplement, Economica*, 1986, *53*, No. 210(S), 1986.

Bergson, Abram, *The Economics of Soviet Planning*, New Haven: Yale University Press, 1964.

Clayton, Elizabeth, "Age-Earnings Profiles of Women in the USSR," *paper presented at the session* "Women in Comparative Perspective" *at the Allied Social Science Associations Convention*, New Orleans, December 28, 1986.

Davies, R. W., "The Ending of Mass Unemployment in the USSR," in David Lane, ed., *Labor and Employment in the USSR*, Brighton: Wheatsheaf Books, 1986, 19–35.

______ **and Wheatcroft, S. G.,** "A Note on the Sources of Unemployment Statistics," in David Lane, ed., *Labor and Employment in the USSR*, Brighton: Wheatsheaf Books, 1986, 36–49.

Ehrenberg, Ronald G., "The Demographic Structure of Unemployment Rates and Labor Market Transition Probabilities," *Research in Labor Economics*, *3*, 1980, 241–92.

Franz, Wolfgang and König, Heinz, "The Nature and Causes of Unemployment in the Federal Republic of Germany since the 1970s: An Empirical Investigation," in Charlie Bean, Richard Layard, and Stephen Nickell, eds., *Unemployment Supplement, Economica*, 1986, *53*, No. 210(S), 1986, S219–S244.

Granick, David, (1987a) *Job Rights in the Soviet Union: Their Consequences*, New York: Cambridge University Press, 1987.

______, (1987b) "Commentary: Consumption and Incentives," in U.S. Congress, Joint Economic Committee, *Gorbachev's Economic Plans*, Vol. 2, Washington: USGPO, November 1987, 336–43.

Gregory, Paul R., "Productivity, Slack, and Time Theft in the Soviet Economy," in James Millar, ed., *Politics, Work, and Daily Life in the USSR*, New York: Cambridge University Press, 1987, 241–78.

______ **and Collier, Irwin L.,** *Soviet Interview Project Working Paper*, No. 35, March 1987.

______ **and Stuart, Robert C.,** *Soviet Economic Structure and Performance*, 3rd ed., New York: Harper & Row, 1986.

Hanson, Philip, "The Serendipitous Soviet Achievement of Full Employment: Labor Shortage and Labor Hoarding in the Soviet Economy," in David Lane, ed., *Labor and Employment in the USSR*, Brighton: Wheatsheaf Books, 1986, 83–111.

Horvath, Francis W., "Forgotten Unemployment: Recall Bias in Retrospective Data," *Monthly Labor Review*, March 1982, *105*, 40–43.

Kotliar, A. E., ed., *Zaniatost' naseleniia*, Moscow: Finansy i Statistika, 1983.

Lampert, Nick, "Job Security and the Law in the USSR," in David Lane, ed., *Labor and Employment in the USSR*, Brighton: Wheatsheaf Books, 1986, 256–77.

Lane, David, ed., *Labor and Employment in the USSR*, Brighton: Wheatsheaf Books, 1986.

Millar, James, ed., *Politics, Work, and Daily Life in the USSR*, New York: Cambridge University Press, 1987.

National Opinion Research Center, *Soviet Interview Project Methodological Report*, submitted to Soviet Interview Project, University of Illinois, October 1985.

Ofer, Gur and Vinokur, Aron, "Earning Differentials by Sex in the Soviet Union," in Steven Rosefielde, ed., *Economic Welfare and the Economics of Soviet Socialism*, Cambridge, U.K.: Cambridge University Press, 1981, 127–62.

______ **and** ______, "The Economics of the Soviet Urban Population in the 1970s: A Report to the National Council for Soviet and East European Research," July 1984.

Organisation for Economic Co-operation and Development (OECD), *Living Conditions in OECD Countries: A Compendium of Social Indicators*, Social Policy Studies, No. 3, Paris: OECD, 1986.

Rosefielde, Steven, ed., *Economic Welfare and the Economics of Soviet Socialism*, Cambridge, U.K.: Cambridge University Press, 1981.

Shapiro, Judith, "Unemployment," *Paper Prepared for the Fourth Conference of the International Work Group on Soviet Economic History*, Birmingham, England, January 1987.

Shipler, David K., *Russia: Broken Idols, Solemn Dreams*, New York: Viking Penguin, 1984.

Sudman, Seymour and Bradburn, Norman M., "Effects of Time and Memory on Response," *Journal of the American Statistical Association*, March 1964, *59*, 18–55.

Swafford, Michael et al., "Response Effects in SIP's General Survey of Soviet Emigrants," in James Millar, ed., *Politics, Work, and Daily Life in the USSR*, New York: Cambridge University Press, 1987, 372–403.

Wiles, Peter, "A Note on Soviet Unemployment on U.S. Definitions," *Soviet Studies*, April 1972, *23*, 619–28.

U.S. Department of Commerce, Bureau of the Census, *Foreign Born Immigrants in U.S. 1980 Census of Housing and Population* (microfiche), Table 4, "Labor Force Participation of Immigrants by Region and Country of Birth: Soviet Union," 4-1741 to 4-1746, 1985.

U.S. Department of Labor, Bureau of Labor Statistics, *Labor Force Statistics from the Current Population Survey: A Databook*, Vol. I, Bulletin 2096, Washington: USGPO, September 1982.

______, *Handbook of Labor Statistics*, Bulletin 2175, Washington: USGPO, December 1983.

______, *Work Experience of the Population in 1969*, Special Labor Force Report 127, Washington: USGPO, 1971.

Costs of Price Adjustment and the Welfare Economics of Inflation and Disinflation

By LEIF DANZIGER*

This paper studies the welfare consequences of inflation and disinflation in a model in which firms incur a fixed cost of price adjustment. It is shown that a moderately increasing price level is associated with higher social welfare than a perfectly stable price level, and the model therefore provides theoretical support for the widespread belief that a little bit of inflation is good for the economy. Nevertheless, it is also true that the eradication of a moderate inflation leads to an increase in social welfare.

In an inflationary environment firms need to adjust their nominal prices in order to offset the decrease in their real prices caused by the increase in the general price level. However, monopolistic firms that incur a fixed cost of price adjustment will not increase their nominal prices continuously, even in the presence of a perfectly anticipated inflation. Instead, the firms will keep their nominal prices constant in periods of discrete length and then increase their prices in discrete jumps.[1]

Within a period with constant nominal price a firm's real price decreases at the rate of inflation, and the real price is therefore first above and then below the profit-maximizing real price the firm would charge if there were no inflation. The decrease in the real price is accompanied by an opposite movement in the output, which is first below and then above the output that would be produced if there were no inflation. Consequently, compared to a situation in which the general price level is stable and the firm charges the profit-maximizing real price at all times, for any given real income the consumers are worse off in the first part of the periods with constant nominal price, but better off in the second part. Furthermore, assuming the real price is always above the real marginal cost of production, it is also true that social welfare is lower in the first part of the periods with constant nominal price, but higher in the second part of the periods.

Whether social welfare is enhanced or impaired by inflation therefore depends on the magnitude of the loss in the first part of the periods relative to the gain in the second

---

*Department of Economics, Bar-Ilan University, Ramat Gan, 52100, Israel. I wish to thank Daniel F. Spulber and two anonymous referees for helpful comments.

[1] The theory of optimal price adjustments when there is a fixed cost of price adjustment was developed by Eitan Sheshinski and Yoram Weiss, 1977, for a perfectly anticipated inflation, and extended by Sheshinski and Weiss, 1983, and Leif Danziger, 1983, 1984, to a stochastic inflation. See also Robert Barro, 1972, and Michael Mussa, 1977. The theory has been applied to study the effects of inflation on the demand for fixed factors (Sheshinski and Weiss, 1979); the aggregate consequences of monetary growth (Mussa, 1981b; Julio Rotemberg, 1983; and Andrew Caplin and Daniel Spulber, 1987); the asymmetric responses to inflation and equivalent deflation (Timur Kuran, 1983); the output effects of inflation (George Akerlof and Janet Yellen, 1985; Kuran, 1986; Howard Naish, 1986; and Michael Parkin, 1986); the possibility of large inefficient fluctuations in real variables (Gregory Mankiw, 1985; Olivier Blanchard and Nobuhiro Kiyotaki, 1987; and Laurence Ball and David Romer, 1987); the possibility of supporting the collusive outcome among oligopolists (Robert Gertner, 1986); and the implications of inflation for the variability and uncertainty of real prices (Stanley Fischer, 1981a; Danziger, 1987a, b; and Roland Benabou, 1986, 1987). For some empirical evidence on the amplitude and frequency of price changes, see Mussa, 1981a; Dennis Carlton, 1986; Stephen Cecchetti, 1986; and Danziger 1987b.

part of the periods, and it is evidently of great theoretical and practical interest to ascertain whether the loss or the gain dominates. In fact, one of the most fundamental questions to be asked of the fixed-cost theory of price adjustment is how the theory predicts that inflation affects the part of social welfare that is derived from the monopolistic firms.

The purpose of this paper is to answer this question, as well as the related question of whether disinflation is socially beneficial.[2] Section I presents a model of an economy with a continuum of monopolistic firms that produce differentiated products and incur a fixed cost when adjusting their nominal price. The general price level increases at a constant rate, and each firm maximizes its discounted real profits by choosing, at each price adjustment, the new nominal price and the length of the period until the next price adjustment. There is a continuum of consumers, and each consumer's utility depends on his consumption of the different products and on his leisure. A consumer earns income by working in the competitive labor market, and he also receives his share of the distributed profits. Each consumer maximizes his utility by choosing how much to work and how to divide his income between the different products.

Section II derives the social welfare, defined as the discounted value of a consumer's utility. It is shown that the utility generated by all the firms is constant and, since the firms have a measure of unity, equal to the average undiscounted utility generated by any one of the firms during a period with constant nominal price.

Section III analyzes the welfare consequences of inflation and proves that social welfare is higher under small positive inflation rates than under full price stability (where each firm charges the profit-maximizing real price). A slowly increasing price level is therefore preferred to a constant price level. The explanation is that discounting of future profits makes a firm's price-adjustment policy asymmetric in the sense that the firm chooses a nominal price for which its real price at the time of price adjustment is only slightly higher than its profit-maximizing real price, but also chooses to keep the nominal price constant for so long that its real prices in most of the period become significantly lower than its profit-maximizing real price. The consumers benefit directly from the lower real prices, and this effect more than outweighs the adverse effect of the lower distributed real profits that the consumers receive qua owners of the firms.

Section IV studies the welfare consequences of disinflation. Ending inflation involves a loss due to the higher real prices that the consumers must pay for some of the products. The explanation is again the asymmetric nature of the price-adjustment policy, which now implies that all firms that adjust their nominal price at the time inflation is brought to an end will do so by increasing their nominal price. However, ending inflation also benefits the firms and hence leads to a gain in the distributed real profits and the consumers' real income. The overall impact of disinflation on social welfare is therefore not immediately evident, but it is proved that the effect of the gain in the real income dominates so that social welfare is always increased when a moderate inflation is brought to an end.

The above analyses of the welfare effects of inflation and disinflation assumed that the inflation rates were relatively low and are probably valid for most of the recent inflation experience, although this of course remains an empirical question. Section V therefore proceeds to examine the welfare implications of inflation and disinflation at more extreme inflation rates, and it presents two examples demonstrating that the welfare effects may be reversed at high inflation rates. Finally, Section VI concludes the paper.

---

[2] It is surprising that the literature has almost ignored the welfare implications of fixed costs of price adjustment. Early exceptions include Mussa, 1977, and Blanchard, 1979. More recently, Akerlof and Yellen, 1985; Mankiw, 1985; Blanchard and Kiyotaki, 1987; Ball and Romer, 1987, and others have shown that small costs of price adjustment may have large efficiency effects

I. The Model

Consider an economy with a continuum of monopolistic firms that produce differentiated nonstorable products. The firms are indexed by i, are uniformly distributed on the interval $(0;1]$, and have a measure of unity. There is also a continuum of consumers that are identical and have a measure of $N>0$. If x_{it} denotes an individual consumer's instantaneous consumption of the ith firm's product at time t, and l_t his instantaneous leisure at time t, then the consumer's instantaneous utility at time t is given by

$$(1) \qquad \int_0^1 u(x_{it})\,di + l_t,$$

where $u(x_{it})$ is increasing, strictly concave, and twice continuously differentiable for $0<x_{it}<\bar{x}$, where $0<\bar{x}\leq\infty$.

A consumer has $I>0$ of time available at each t. He works some of his time for the wage rate in the competitive labor market, and enjoys the remainder as leisure. All consumers own an equal fraction of all firms, and profits are immediately distributed. Using labor as the numeraire, real prices and real profits are defined as nominal prices and nominal profits divided by the wage rate. If z_{it} denotes the ith firm's real price at time t, and A_t the instantaneous real profits distributed to each consumer at time t, then a consumer's instantaneous budget constraint at time t is given by

$$(2) \qquad I + A_t = \int_0^1 z_{it}x_{it}\,di + l_t.$$

The objective of each consumer is to choose his instantaneous consumption and leisure in order to maximize his instantaneous utility (1) given his instantaneous budget constraint (2). Assuming that a solution with $0<x_{it}<\bar{x}$ and $l_t>0$ exists, this implies that $\partial u(x_{it})/\partial x_{it} = z_{it}$ for all i; consequently, a consumer's demand (density) for the ith firm's product depends on only z_{it} and may be written as $D(z_{it})$. The aggregate demand for the ith firm's product is therefore $ND(z_{it})$.

Labor is the only factor of production and is freely variable.[3] All firms require a labor input of $k\geq 0$ for each unit of output, and if the real price exceeds k, the ith firm's (maximum) instantaneous real profit from production at time t is given by[4]

$$\Pi(z_{it}) \equiv (z_{it}-k)ND(z_{it}).$$

It is assumed that the instantaneous real profit from production is maximized at the real price Z, $0<Z<\infty$, and strictly quasi-concave when positive.

The general price level (as measured by the wage rate) increases at the constant rate $\mu>0$.[5] Due to the fixed cost of price adjustment, however, firms do not adjust their nominal prices continuously. Instead, each firm keeps its nominal price constant in periods of discrete length, after which it increases its price by a discrete jump. Thus, if S denotes the initial real price, that is, the real price at the beginning of a period with constant nominal price, then $Se^{-\mu\tau}$ is the real price after $\tau\geq 0$ of the period has elapsed. Let T denote the duration of a period with constant nominal price (i.e., if the nominal price is adjusted at $t=0$, then it is kept constant in $[0;T)$), $C>0$ the fixed real cost of price adjustment, and $r>0$ the real interest rate. At the time of price adjustment, the firm's discounted real profits from the period with constant nominal price become

$$\int_0^T e^{-r\tau}\Pi(Se^{-\mu\tau})\,d\tau - C.$$

---

[3] Inclusion of other factors (variable or fixed) would be inconsequential. The results would also hold with labor as a fixed factor if there were another variable factor.

[4] Of course, if the real price does not exceed k, the (maximum) instantaneous real profit is zero. The analysis below indicates that unless the real price always exceeds k, it would be optimal for the firm to cease to exist.

[5] In fact, with the assumptions of the model, any symmetric and homogeneous aggregation of the different firms' prices also increases at rate μ. See Rotemberg, 1983, and Caplin and Spulber, 1987.

It may be assumed that the firm chooses the same S and T at each price adjustment (see fn. 7 below). At the time of a price adjustment, the total discounted real profits therefore become

$$V \equiv (1-e^{-rT})^{-1}\left[\int_0^T e^{-r\tau}\Pi(Se^{-\mu\tau})\,d\tau - C\right].$$

Each firm determines S and T so as to maximize V. The first-order conditions for a maximum are

$$(3)\qquad \frac{\partial V}{\partial S} = (1-e^{-rT})^{-1}\int_0^T e^{-(\mu+r)\tau} \times \Pi'(Se^{-\mu\tau})\,d\tau = 0,$$

$$(4)\qquad \frac{\partial V}{\partial T} = (e^{rT}-1)^{-1}[\Pi(s) - rV] = 0,$$

where $s \equiv Se^{-\mu T}$ is the terminal real price, that is, the real price at the end of a period with constant nominal price.[6] Condition (3) shows that S is determined so that the discounted value of the marginal real profits from a change in S is equal to zero, while condition (4) shows that T is determined so that the real profit at the terminal real price is equal to the "permanent" real profit rV. Assuming that $V > 0$ (otherwise it would be optimal for a firm to cease to exist), it follows that $\Pi(s) > 0$ and hence $s > k$, so that a firm's real price always exceeds k.

Using the strict quasiconcavity of $\Pi(z_{it}) > 0$, it may be shown that whenever conditions (3) and (4) are satisfied, then $s < Z < S$, and the second-order conditions for a maximum are also satisfied. It follows that the (S,T) policy is unique[7] and will be followed by all firms.

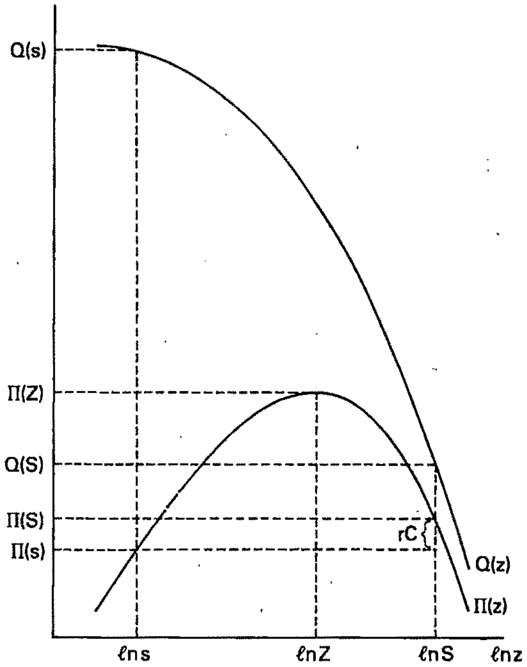


FIGURE 1

By partial integration and the definition of V,

$$\int_0^T e^{-(\mu+r)\tau}\Pi'(Se^{-\mu\tau})\,d\tau$$

$$= -\left.\frac{e^{-r\tau}\Pi(Se^{-\mu\tau})}{\mu S}\right|_0^T - \frac{r}{\mu S}\int_0^T e^{-r\tau}\Pi(Se^{-\mu\tau})\,d\tau$$

$$= \frac{1}{\mu S}\left[\Pi(S) - e^{-rT}\Pi(s) - rV(1-e^{-rT}) - rC\right],$$

which, by use of condition (4), is equal to

$$\frac{1}{\mu S}\left[\Pi(S) - \Pi(s) - rC\right].$$

Conditions (3) and (4) are therefore equivalent to

---

[6] More precisely, s is the limit of the real prices as τ approaches T.

[7] See Sheshinski and Weiss, 1977, on which the derivation of the individual firm's price-adjustment policy is based. The uniqueness of the (S,T) policy provides a justification for assuming that a firm chooses the same S and T at each price adjustment.

$$\Pi(S)-\Pi(s)-rC=0, \tag{5}$$

$$\Pi(s)-rV=0. \tag{6}$$

Condition (5) expresses that the real profit at the initial real price is equal to the real profit at the terminal real price plus the interest saved by postponing a price adjustment. This condition is illustrated by the $\Pi(z)$ curve in Figure 1 (note that the horizontal axis measures the logarithm of the real price).

II. The Social Welfare

Let t_i denote the time of the ith firm's last price adjustment before t, and assume that t_i can be expressed as $t_i=t-iT$. Since i is uniformly distributed on (0;1], it follows that t_i is uniformly distributed on $[t-T;t)$; that is, the firms are uniformly distributed with density $1/T$ in terms of the time of their last price adjustment before t. The real cost of price adjustment incurred at time t is, therefore, C/T, and the total real profits generated by all firms' production at time t less the real cost of price adjustment at time t become

$$\int_0^1\Pi(z_{it})\,di-\frac{C}{T}.$$

The instantaneous real profits distributed to each consumer at time t are then

$$A_t=\left[\int_0^1\Pi(z_{it})\,di-\frac{C}{T}\right]\frac{1}{N}.$$

The consumers' expenditures are equal to the firms' revenues, so substitution of A_t in the budget constraint (2) shows that a consumer's instantaneous leisure at time t is

$$\begin{aligned} l_t&=I+\left[\int_0^1\Pi(z_{it})\,di-\frac{C}{T}\right]\frac{1}{N}\\ &\quad-\int_0^1 z_{it}D(z_{it})\,di\\ &=I-k\int_0^1 D(z_{it})\,di-\frac{c}{T}, \end{aligned}$$

where c is defined by $c\equiv C/N$. A consumer's (maximized) instantaneous utility at time t therefore becomes

$$\begin{aligned} U_t&\equiv\int_0^1 u[D(z_{it})]\,di+I-k\int_0^1 D(z_{it})\,di-\frac{c}{T}\\ &=\int_0^1 Q(z_{it})\,di+I-\frac{c}{T}, \end{aligned}$$

where $Q(z_{it})\equiv u[D(z_{it})]-kD(z_{it})$ is the gain in utility derived from the ith firm's production of $D(z_{it})$. This gain of course decreases with z_{it} for $s<z_{it}\leq S$, since the higher z_{it} is, the fewer units are bought relative to the $D(k)$ units for which the gain is maximized.[8]

It follows from the definition of s that $T=(\ln S-\ln s)/\mu$, and from $t_i=t-iT$ that $z_{it}=Se^{-\mu iT}\Rightarrow i=(\ln S-\ln z_{it})/(\mu T)=(\ln S-\ln z_{it})/(\ln S-\ln s)$. So by changing the integration variable from i to $\ln z$ and substituting for T, one obtains that

$$U_t=\frac{\int_s^S Q(z)\,d\ln z-\mu c}{\ln S-\ln s}+I.$$

Accordingly, the instantaneous utility is constant. Furthermore, since the firms have a measure of unity, the instantaneous utility generated by all the firms is equal to the average undiscounted instantaneous utility generated by any one of the firms during one period with constant nominal price. Thus, in Figure 1 the instantaneous utility generated by all the firms is, except for the cost of

---

[8] Formally,

$$\begin{aligned} \frac{dQ(z_{it})}{dz_{it}}&=\left\{\frac{\partial u[D(z_{it})]}{\partial D(z_{it})}-k\right\}\frac{dD(z_{it})}{dz_{it}}\\ &=(z_{it}-k)\frac{dD(z_{it})}{dz_{it}}<0, \end{aligned}$$

noting that $\partial u[D(z_{it})]/\partial D(z_{it})=z_{it}$, $z_{it}-k>0$, and $dD(z_{it})/dz_{it}<0$ from the second-order condition for maximum.

price adjustment, equal to the average value of $Q(z)$ as $\ln z$ decreases from $\ln S$ to $\ln s$.

Social welfare is defined as the discounted (at rate r) value of $U_t, t \geq 0$. It becomes

$$W \equiv \int_0^{\infty} e^{-rt} U_t \, dt$$

$$(7) \qquad = \left[\frac{\int_s^S Q(z)\, d \ln z - \mu c}{\ln S - \ln s} + I \right] \frac{1}{r}.$$

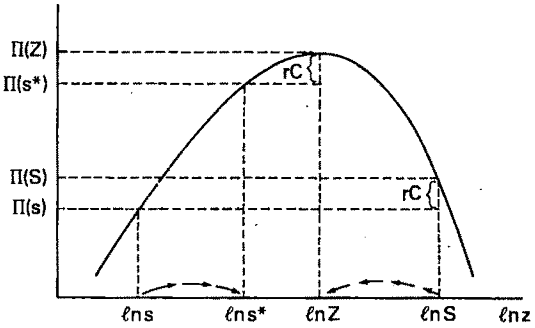


FIGURE 2

III. The Welfare Economics of Inflation

If there were no inflation, then each firm would maximize its total discounted real profits by choosing $S = Z$ and $T = \infty$ (while s would be indeterminate). The discounted real profits would become $\Pi(Z)\int_0^{\infty} e^{-rt} dt - C = \Pi(Z)/r - C$, which is also the limiting value of V as μ approaches zero. It follows from conditions (5) and (6) that $\lim_{\mu \to 0} \Pi(S) = \Pi(Z)$ and $\lim_{\mu \to 0} \Pi(s) = \Pi(Z) - rC$. Consequently, in view of the strict quasiconcavity of Π for $\Pi > 0$, the limiting values of S and s as the inflation rate approaches zero are Z and s^*, where s^* is defined by $s^* \equiv \Pi^{-1}[\Pi(Z) - rC]$ and $s^* < Z$. Inflation in the limit simply erodes monopoly power. This is illustrated in Figure 2 where the arrows indicate the movements of $\ln S$ and $\ln s$ toward $\ln Z$ and $\ln s^*$ as μ approaches zero. The important point to note is that, because of discounting, the terminal real price does *not* approach Z as μ approaches zero: the limiting distribution of the real prices as μ approaches zero is concentrated on the interval $(s^*; Z]$, almost all of which is *below* the real price Z that would be charged in the absence of inflation.[9] In other words, when the rate of inflation is low, then the discounting of future profits introduces an asymmetry in a firm's price-adjustment policy: on the one hand, the firm chooses a nominal price for which its real price at the time of price adjustment is only slightly higher than its profit-maximizing real price, and on the other hand, the firm also chooses to keep its nominal price constant for so long a period that its real prices in most of the period become significantly lower than its profit-maximizing real price.

The limiting value of the social welfare as the inflation rate approaches zero, becomes

$$W^* \equiv \lim_{\mu \to 0} W$$

$$= \left[\frac{\int_{s^*}^{Z} Q(z)\, d \ln z}{\ln Z - \ln s^*} + I \right] \frac{1}{r}.$$

Since $Q(z)$ is a decreasing function of z for $s^* < z \leq Z$, the limiting distribution of the contributions of the different firms to the social welfare is concentrated on the interval $[Q(Z); Q(s^*))$, almost all of which is *above* the contribution $Q(Z)$ that corresponds to the real price Z. In other words, since the consumers' gain from lower real prices outweigh their loss due to lower distributed real profits, it follows that W^* exceeds what the social welfare would be in the absence of inflation. Because of the continuity of S and s in the inflation rate, also W is continuous in the inflation rate. The social welfare is therefore higher under small positive inflation rates than in the absence of inflation, and it may be concluded that a steady-state moderately increasing price level is preferable to perfect price stability. Accordingly, the model provides theoretical support for the widely held belief that the economy benefits from a little bit of inflation.

---

[9] This reflects the fact that if there were no inflation, then the discounted real profits from any real price within the interval (s^*, Z) exceed the discounted real profits from incurring the cost of price adjustment and obtaining the real price Z. That is, $\Pi(z)/r > \Pi(Z)/r - C$ for $s^* < z < Z$.

Since the result that social welfare is higher under moderate inflation rates than in the absence of inflation is caused by the fact that under moderate inflation rates most real prices are below the real price that would be charged in the absence of inflation, it provides no clue as to how social welfare varies with a (positive) steady-state inflation rate. To determine this, totally differentiate W with respect to μ,

$$\frac{dW}{d\mu}=\left\{\left[\frac{Q(S)+I-rW}{S}\right]\frac{dS}{d\mu}-\left[\frac{Q(s)+I-rW}{s}\right]\frac{ds}{d\mu}-c\right\}\times r^{-1}(\ln S-\ln s)^{-1}.$$

Also, totally differentiate conditions (5) and (6) with respect to μ,

$$\frac{dS}{d\mu}=\frac{r\partial V/\partial\mu}{\Pi'(S)},\qquad\qquad (8)$$

$$\frac{ds}{d\mu}=\frac{r\partial V/\partial\mu}{\Pi'(s)}.$$

Hence,

$$\frac{dW}{d\mu}=\left\{\left[\frac{Q(S)+I-rW}{S\Pi'(S)}-\frac{Q(s)+I-rW}{s\Pi'(s)}\right]\frac{\partial V}{\partial\mu}-\frac{c}{r}\right\}\times(\ln S-\ln s)^{-1}.$$

Observe that

$$\lim_{\mu\to 0}[Q(S)+I-rW]=Q(Z)+I-rW^*<0,$$

$$\lim_{\mu\to 0}[Q(s)+I-rW]=Q(s^*)+I-rW^*<\infty,$$

and, by the continuous differentiability of $\Pi(z)$ (because $u(x_{it})$ is twice continuously differentiable in x_{it}), that

$$\lim_{\mu\to 0}[S\Pi'(S)]=0_-,$$

$$\lim_{\mu\to 0}[s\Pi'(s)]=s^*\Pi'(s^*)>0.$$

Therefore,

$$\lim_{\mu\to 0}\left[\frac{Q(S)+I-rW}{S\Pi'(S)}-\frac{Q(s)+I-rW}{s\Pi'(s)}\right]=\infty.$$

Since condition (3) and the strict quasiconcavity of Π for $\Pi>0$ implies that

$$\frac{\partial V}{\partial\mu}=-(1-e^{-rT})^{-1}S\times\int_0^T\tau e^{-(\mu+r)\tau}\Pi'(Se^{-\mu\tau})\,d\tau<0,\qquad (9)$$

it follows that $\lim_{\mu\to 0}dW/d\mu<0$. It may therefore be concluded that the social welfare decreases with the inflation rate in a neighborhood of $\mu=0$. This is illustrated in Figure 3: for low inflation rates the social welfare is inversely related to inflation rate, even though the social welfare is always above the no-inflation level. The reason is that the profit function is very flat for real prices close to Z, so for low inflation rates almost all of the effect of a higher μ will be through a higher S (which implies a lower W), while only very little will be through a lower s (which implies a higher W).

It is also of interest to determine how the social welfare depends on the fixed cost of price adjustment. Since W is continuous in the inflation rate, $\lim_{\mu\to 0}dW/dC=dW^*/dC$, so by totally differentiating W^* with respect to C,

$$\lim_{\mu\to 0}\frac{dW}{dC}=\left[\frac{rW^*-I-Q(s^*)}{s^*}\right]\frac{ds^*}{dC}\times r^{-1}(\ln Z-\ln s^*)^{-1}.$$

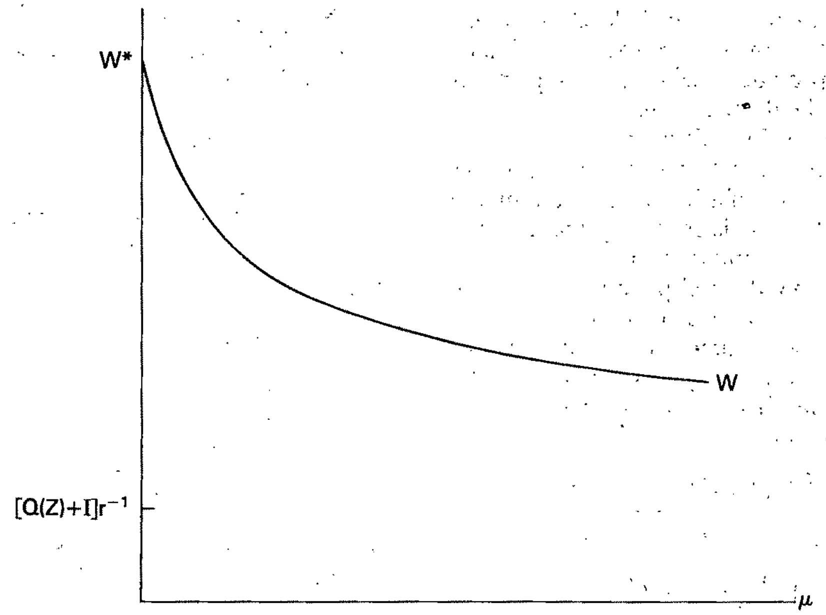


FIGURE 3

Now,

$$\frac{rW^* - I - Q(s^*)}{s^*} < 0,$$

and, by the definition of s^*,

$$\frac{ds^*}{dC} = -\frac{r}{\Pi'(s^*)} < 0.$$

Accordingly, $\lim_{\mu \to 0} dW/dC > 0$. So, one obtains the rather paradoxical result that for low inflation rates social welfare increases with the fixed cost of price adjustment (i.e., for low inflation rates the *W* curve in Figure 3 shifts upward). The logic is, however, straightforward: the fixed cost of price adjustment reduces the terminal real price, and since the adjustment cost is never incurred in the limit as the inflation rate approaches zero, at low rates of inflation the indirect positive effect of the downward push in real prices dominates the direct negative effect of the higher cost of price adjustment.

IV. The Welfare Economics of Disinflation

The reduction of inflation is one of the major goals of economic policy, and a question of great practical significance is how a *transition* from inflation to price stability affects welfare. An answer to this question cannot, however, be obtained from the above analysis which only compared *steady states* with different levels of inflation. In order to assess the welfare effects of disinflation, assume now that at some point in time inflation is permanently eradicated.[10] Some firms will then find it profitable to immediately adjust their nominal price so as to make their real price equal to *Z*, whereas other firms will find it profitable to keep their nominal price unchanged forever. Since the discounted real profits from immediately adjusting the nominal price are $\Pi(Z)\int_0^\infty e^{-rt}\,dt$

---

[10] It is also assumed that disinflation is totally unanticipated and hence does not affect the firms' price setting before the disinflation occurs (this would, anyway, be inconsequential for the low inflation rates considered in this section).

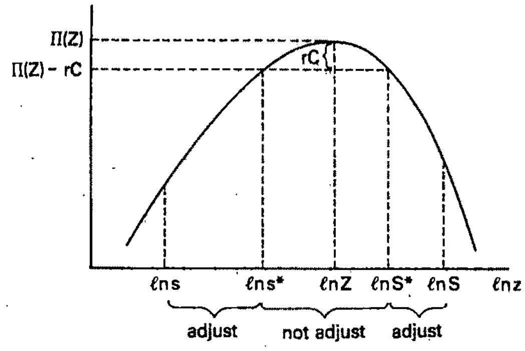


FIGURE 4

$-C = \Pi(Z)/r - C$, while the discounted real profits from keeping the nominal price (and the corresponding real price z) unchanged are $\Pi(z)\int_0^{\infty} e^{-rt}\,dt = \Pi(z)/r$, firms for which $\Pi(z)/r < \Pi(Z)/r - C$ will adjust their nominal price, while firms for which $\Pi(z)/r > \Pi(Z)/r - C$ will keep their nominal price unchanged. Let S^* be defined by $S^* \equiv \Pi^{-1}[\Pi(Z) - rC]$ and $S^* > Z$; that is, $\Pi(S^*) = \Pi(s^*)$ and $s^* < Z < S^*$. Consequently, any firm whose real price either falls below s^* or above S^* at the time inflation is brought to an end will choose to adjust its nominal price so that its real price becomes Z, while any firm whose real price falls between s^* and S^* at the time inflation is brought to an end will choose not to adjust its nominal price.[11] This is illustrated in Figure 4. It is clear that $ds^*/dr < 0$ and $dS^*/dr > 0$; that is, the higher the interest rate, the fewer firms find it profitable to incur the price-adjustment cost at the time of disinflation in order to make the real price equal to Z.

There always exist firms whose real prices under inflation fall below s^* (since $ds/d\mu < 0$). For these firms disinflation leads to a higher real price and a lower production. On the other hand, because of the asymmetric nature of the price-adjustment policy, for low inflation rates there will not be any firm whose real price is above S^* (since $Z < S^*$), and therefore no firm whose real price is so high that disinflation leads to a lower real price and a higher production. The eradication of a moderate inflation therefore necessarily involves a loss due to the higher real prices that the consumers must pay for the products of some of the firms. At the same time, disinflation also leads to an increase in the firms' real profits,[12] and hence to a gain due to the higher distributed real profits that cause the consumers' real income to increase. Because of these opposite effects, the overall impact on social welfare is not immediately evident.

It will now be proved, however, that bringing a moderate inflation to an end always increases social welfare. Assume that the inflation rate is sufficiently low that $S \le S^*$. At the time of disinflation, there is a one-time total price-adjustment cost of $C(\ln s^* - \ln s)/(\ln S - \ln s)$, so the social welfare from

---

[11] The firm whose real price is equal to s^* or S^* at the time inflation is brought to an end will be indifferent between adjusting its nominal price and leaving it unchanged.

[12] Under inflation, the discounted real profits of the ith firm at time t are

$$V_i \equiv \int_0^{T_i} e^{-r\tau}\Pi[Se^{-\mu(t-t_i+\tau)}]\,d\tau + e^{-rT_i}V,$$

where $T_i \equiv t_i + T - t$. Let $\hat{T}$ be defined by $\Pi(Se^{-\mu\hat{T}}) \equiv \Pi(S)$ and $\hat{T} > 0$. Since the choice of (S, T) maximizes V, it follows that $V_i < V + C$. Consequently, if $t \le t_i + \hat{T}$, conditions (5) and (6) imply that

$$V_i < \Pi(s)/r + C = \Pi(S)/r \le \Pi[Se^{-\mu(t-t_i)}]/r.$$

Since $\Pi[Se^{-\mu(t-t_i)}]/r$ are the discounted real profits under disinflation if the firm does not adjust its nominal price, this shows that disinflation increases the discounted real profits if $t \le t_i + \hat{T}$. On the other hand, if $t > t_i + \hat{T}$, the fact that $\Pi(Se^{-\mu\hat{T}}) > \Pi(Se^{-\mu(t-t_i)}]$ together with conditions (5) and (6) imply that

$$\begin{aligned} V_i &< \Pi(Se^{-\mu\hat{T}})\int_0^{T_i} e^{-r\tau}\,d\tau + e^{-rT_i}V \\ &= \Pi(S)(1 - e^{-rT_i})/r + \Pi(s)e^{-rT_i}/r \\ &< \Pi(S)/r = \Pi(s)/r + C < \Pi(Z)/r + C. \end{aligned}$$

Since $\Pi(Z)/r + C$ are the discounted real profits if the firm adjusts its nominal price, this shows that disinflation also increases the discounted real profits if $t > t_i + \hat{T}$.

disinflation becomes

$$\left[\frac{Q(Z)\int_{s}^{s^*} d\ln z+\int_{s^*}^{S} Q(z)\,d\ln z-rc(\ln s^*-\ln s)}{\ln S-\ln s}+I\right]\frac{1}{r}.$$

Accordingly, the change in the social welfare is (by (7)),

$$G\equiv\left\{\frac{\int_{s}^{s^*}[Q(Z)-Q(z)]\,d\ln z+\mu c-rc(\ln s^*-\ln s)}{\ln S-\ln s}\right\}\frac{1}{r}$$

$$=\left\{\frac{\int_{s}^{s^*}[Q(Z)-Q(z)-rc]\,d\ln z+\mu c}{\ln S-\ln s}\right\}\frac{1}{r},$$

which has the same sign as

$$H\equiv\int_{s}^{s^*}[Q(Z)-Q(z)-rc]\,d\ln z+\mu c.$$

To determine the sign of H, totally differentiate H with respect to μ,

$$\frac{dH}{d\mu}=[-Q(Z)+Q(s)+rc]\frac{ds/d\mu}{s}+c.$$

Next, substitute $\partial V/\partial\mu$ from equation (9) into equation (8) to obtain that

$$\frac{ds}{d\mu}=-\frac{rS\int_{0}^{T}\tau e^{-(\mu+r)\tau}\Pi'(Se^{-\mu\tau})\,d\tau}{\Pi'(s)(1-e^{-rT})}.$$

Since

$$\lim_{\tau\to\infty}[\tau e^{-(\mu+r)\tau}]=0,$$

and

$$\lim_{\mu\to 0}\Pi'(Se^{-\mu\tau})=0$$

for any finite τ, it follows that $\lim_{\mu\to 0}ds/d\mu=0$. Therefore, it may be concluded that $dH/d\mu>0$ for small μ, and since $\lim_{\mu\to 0}H=0$, also that $H>0$ for small μ.

Consequently, disinflation has a positive impact on social welfare. So even though disinflation causes higher real prices for the products of some of the firms, due to the positive effect on the firms' real profits, the eradication of a slow inflation is socially beneficial. Intuitively, the explanation is that for moderate inflation rates s is a very flat function of the inflation rate so that only few firms adjust their nominal price at the time of disinflation; the positive effect of avoiding the costs of price adjustment in the whole future therefore dominates the negative effect of the higher real prices and the extra costs of price adjustment incurred at the time of disinflation.

V. Two Examples

In many practical applications it is probably reasonable to assume that the inflation rates are sufficiently low that they lie within the range for which the results in Sections III and IV were obtained. Nonetheless, one would like to explore whether these results hold globally or may be reversed for high inflation rates. For instance, is it possible that the social welfare decreases so much with the inflation rate that it eventually falls below the social welfare that would be obtained in the absence of inflation? Or conversely, is it possible that the social welfare at some point starts to increase with the inflation rate and eventually rises above W^*? And is it possible that for high inflation rates disinflation may have a negative effect on the social welfare? Since the proofs of the results in Sections III and IV hinge on the asymmetry in the firms' price-adjustment policy caused by the interaction between a positive discounting and a low inflation rate, it is perhaps not surprising that the results may indeed be reversed for high inflation rates. This will be demonstrated with the help of two examples. In the first example the demand function is loglinear, and it is shown that at high inflation rates the social welfare is below its no-inflation level and decreases with the fixed cost of price adjustment. In the second example the demand function is linear, and it is shown that at high inflation rates the social welfare is above W^*, increases with the inflation rate, and is negatively effected by disinflation.

A. *The First Example*

Assume that $u(x_{it}) = 2x_{it}^{\frac{1}{2}}$ and $k = 1$. A consumer's demand for the ith firm's product is then $D(z_{it}) = z_{it}^{-2}$, and the instantaneous real profit from production becomes $(z_{it}^{-1} - z_{it}^{-2})N$. Accordingly,

$$V = (1 - e^{-rT})^{-1} \times \left[N \int_0^T e^{-r\tau}(S^{-1}e^{\mu\tau} - S^{-2}e^{2\mu\tau})\, d\tau - C \right].$$

If it is assumed that $\mu = r$, then condition (3) implies that S satisfies

$$S = \frac{2}{T}\int_0^T e^{\mu\tau}\, d\tau.$$

Since the instantaneous real profit distributed to each consumer is constant and given by

$$A = \frac{\int_s^S [zD(z) - D(z)]\, d\ln z - \mu c}{\ln S - \ln s},$$

the social welfare may be written as

$$W = \left(\frac{\int_s^S \{u[D(z)] - D(z)\}\, d\ln z - \mu c}{\ln S - \ln s} + I \right) \frac{1}{r}$$

$$= \left(\frac{\int_s^S \{u[D(z)] - zD(z)\}\, d\ln z}{\ln S - \ln s} + A + I \right) \frac{1}{r}$$

$$= \left[\frac{\int_s^S (2z^{-1} - z^{-1})\, d\ln z}{\ln S - \ln s} + A + I \right] \frac{1}{r}$$

$$= \left(\frac{s^{-1} - S^{-1}}{\ln S - \ln s} + A + I \right) \frac{1}{r}.$$

Substituting for S and using that $\ln S - \ln s = \mu T$,

$$W = \left(\tfrac{1}{2} + A + I\right) r^{-1}.$$

Clearly, $Z = 2$ so that $u[D(Z)] - ZD(Z) = \frac{1}{2}$, and $A < \Pi(Z)$, since by definition Z is the profit-maximizing real price. Consequently, there exist high inflation rates for which the social welfare under inflation is less than what it would be in the absence of inflation. Furthermore, since $\lim_{C \to 0} A = \Pi(Z)$,[13] it also follows that $dA/dC < 0$ is possible and hence that $dW/dC < 0$ is possible. So at high inflation rates social welfare may decrease with the fixed cost of price adjustment.

B. *The Second Example*

Assume that $u(x_{it}) = x_{it} - \frac{1}{2}x_{it}^2$ for $x_{it} < 1$ and that $k = 0$. A consumer's demand for the ith firm's product is then $D(z_{it}) = 1 - z_{it}$, and the instantaneous real profit from production becomes $(z_{it} - z_{it}^2)N$. Thus,

$$V = (1 - e^{-rT})^{-1} \times \left[N \int_0^T e^{-r\tau}(Se^{-\mu\tau} - S^2 e^{-2\mu\tau})\, d\tau - C \right].$$

---

[13] Totally differentiating conditions (5) and (6) with respect to C yields

$$\frac{dS}{dC} = -\frac{r}{\Pi'(S)(e^{rT} - 1)},$$

$$\frac{ds}{dC} = -\frac{r}{\Pi'(s)(1 - e^{-rT})},$$

while conditions (5) and (6) together with the strict quasiconcavity of Π for $\Pi > 0$ show that $\lim_{C \to 0} S = \lim_{C \to 0} s = Z$. Therefore,

$$\lim_{C \to 0} T = \frac{1}{\mu}\left(\lim_{C \to 0} \ln S - \lim_{C \to 0} \ln s \right) = 0,$$

$$\lim_{C \to 0} \frac{dS}{dC} = \infty,$$

and

$$\lim_{C \to 0} \frac{dS/dC}{ds/dC} = -1.$$

By l'Hospital's rule,

$$\lim_{C \to 0} A = \frac{D(Z) - D(Z)/Z}{1/Z} = \Pi(Z).$$

By condition (3), S satisfies

$$S=\frac{\int_0^T e^{-(\mu+r)\tau}\,d\tau}{2\int_0^T e^{-(2\mu+r)\tau}\,d\tau},$$

and the discounted real profits from a period with constant nominal price become

$$\frac{N\left[\int_0^T e^{-(\mu+r)\tau}\,d\tau\right]^2}{4\int_0^T e^{-(2\mu+r)\tau}\,d\tau}-C. \tag{10}$$

Let $\bar{\mu}\equiv[1-4rc+(1-4rc)^{1/2}]/(4c)$. It is easy to see that (10) decreases in μ (because of (9)), increases in T, is negative for $\mu\geq\bar{\mu}$ and $T<\infty$, and vanishes for $\mu=\bar{\mu}$ and $T=\infty$. It may therefore be concluded that T approaches infinity as μ approaches $\bar{\mu}$.

The social welfare becomes

$$W=\left\{\frac{\int_s^S u[D(z)]\,d\ln z-\mu c}{\ln S-\ln s}+I\right\}\frac{1}{r}$$

$$=\left[\frac{\int_s^S(1-z^2)\,d\ln z-\mu c}{2(\ln S-\ln s)}+I\right]\frac{1}{r}$$

$$=\left[\tfrac{1}{2}-\frac{(S^2-s^2)/2+\mu c}{2(\ln S-\ln s)}+I\right]\frac{1}{r}$$

$$=\left\{\tfrac{1}{2}-\frac{\left[\int_0^T e^{-(\mu+r)\tau}\,d\tau\right]^2\int_0^T e^{-2\mu\tau}\,d\tau}{8T\left[\int_0^T e^{-(2\mu+r)\tau}\,d\tau\right]^2}\right.$$

$$\left.-\frac{c}{2T}+I\right\}\frac{1}{r}.$$

Substitution of $\mu=\bar{\mu}$ and $T=\infty$ shows that W approaches $(\frac{1}{2}+I)/r$ as μ approaches $\bar{\mu}$. Since $W^*=\lim_{\mu\to 0}W<[u(1)+I]/r=(\frac{1}{2}+I)/r$, one obtains that $W>W^*$ for μ high enough. That is, as μ approaches $\bar{\mu}$, the period with constant nominal price becomes so long that the real prices are close to zero almost all the time. Since zero is the socially optimal price, it may be concluded that there exist high inflation rates for which the social welfare increases with the inflation rate and eventually exceeds W^*. Furthermore, from the fact that $\lim_{\mu\to\bar{\mu}}W=(\frac{1}{2}+I)/r$, it may also be concluded that for high enough inflation rates, disinflation leads to a decrease in social welfare.

VI. Conclusion

This paper has studied the welfare consequences of inflation and disinflation in a model with monopolistic firms that produce differentiated products and incur a fixed cost of price adjustment. A major result is that a moderately increasing price level is associated with a higher level of social welfare than a perfectly stable price level, and the model therefore provides theoretical support for the widespread belief that a little bit of inflation is good for the economy. Another important result is that the eradication of a moderate inflation leads to an increase in social welfare. There is no contradiction between these results: the first result follows from the fact that at low inflation rates a firm's real price is most of the time below the real price that it would charge in the absence of inflation, while the second result follows from the avoidance of costly price adjustments after inflation has been brought to an end.

The analysis has assumed that all consumers are identical and own the same fraction of all firms. Neither inflation nor disinflation, therefore, have distributional effects. In a more realistic setting, however, consumers may be different and ownership of the firms not evenly distributed, so that distributional aspects cannot be ignored. For instance, under inflation the individuals qua consumers benefit from the relatively low real prices, but qua owners of the firms suffer from the relatively low distributed real profits, while after disinflation it is just the opposite. Accordingly, if ownership of the firms were not evenly distributed, but concentrated in the hands of a few individuals, then these individuals would be made worse off by inflation and better off by disinflation, while the other individuals would be made better off by inflation and worse off by disinflation. Thus, the different individuals would disagree about the desirability of disinflation, and this might help explain why anti-inflation policies are so hard to implement.[14]

The existence of price-adjustment costs is of course not the only reason for inflation to have welfare effects. As is well known, the traditional analysis of the welfare consequences of inflation focuses on the socially inefficient economizing on cash balances associated with the nonpayment of interest on cash,[15] and from this viewpoint the optimal rate of inflation is negative and equal to minus the real rate of interest.[16] The present analysis implies that deflation decreases welfare (for the same reason that inflation increases welfare), so it would be interesting to combine the two approaches in order to determine under which conditions there exists a welfare-maximizing rate of inflation, taking into account both the effects of the cost of price adjustment and the effects of the inefficient economizing on cash balances.

---

[14] It has also been assumed that inflation is fully anticipated. This is, however, not essential. It is straightforward to show that the results would also hold with a stochastic inflation, either of the discrete type in Danziger, 1984, or of the continuous type in Caplin and Spulber, 1987.

[15] See Martin Bailey, 1956, for the original statement and Fischer, 1981b, for a more recent application.

[16] See Milton Friedman, 1969.

REFERENCES

Akerlof, George A. and Yellen, Janet L., "A Near-Rational Model of the Business Cycle, with Wage and Price Inertia," *Quarterly Journal of Economics*, Supplement 1985, *100*, 823–38.

Bailey, Martin, "The Welfare Cost of Inflationary Finance," *Journal of Political Economy*, June 1956, *61*, 93–110.

Ball, Laurence and Romer, David, "The Equilibrium and Optimal Timing of Price Changes," mimeo., 1987.

Barro, Robert J., "A Theory of Monopolistic Price Adjustment," *Review of Economic Studies*, January 1972, *39*, 17–26.

Benabou, Roland, "Search, Price Setting and Inflation," mimeo., 1986.

______, "Optimal Price Dynamics and Speculation with a Storable Good," mimeo., 1987.

Blanchard, Olivier Jean, "Wage Indexing Rules and the Behavior of the Economy," *Journal of Political Economy*, August 1979, *87*, 798–815.

______, **and Kiyotaki, Nobuhiro**, "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review*, September 1987, *77*, 647–66.

Caplin, Andrew S. and Spulber, Daniel F., "Menu Costs and the Neutrality of Money," *Quarterly Journal of Economics*, November 1987, *102*, 703–25.

Carlton, Dennis W., "The Rigidity of Prices," *American Economic Review*, September 1986, *76*, 637–58.

Cecchetti, Stephen G., "The Frequency of Price Adjustment: A Study of the Newsstand Prices of Magazines," *Journal of Econometrics*, April 1986, *31*, 255–74.

Danziger, Leif, "Price Adjustments with Stochastic Inflation," *International Economic Review*, October 1983, *24*, 699–707.

______, "Stochastic Inflation and the Optimal Policy of Price Adjustment," *Economic Inquiry*, January 1984, *22*, 98–108.

______, (1987a) "On Inflation and Real Price Variability," *Economic Inquiry*, April 1987, *25*, 285–98.

______, (1987b) "Inflation, Fixed Cost of Price Adjustment, and the Measurement of Relative Price Variability: Theory and Evidence," *American Economic Review*, September 1987, *77*, 704–13.

Fischer, Stanley, (1981a) "Relative Shocks, Relative Price Variability, and Inflation," *Brookings Papers on Economic Activity*, 2:1981, 321–441.

______, (1981b) "Towards an Understanding of the Costs of Inflation: II," in Karl Brunner and Allan H. Meltzer, eds., *Carnegie-Rochester Conference Series on Public Policy: The Costs and Consequences of Inflation*, Autumn 1981, *15*, 5–41.

Friedman, Milton, "The Optimum Quantity of Money," in Milton Friedman, ed., *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine, 1969.

Gertner, Robert, "Dynamic Duopoly with Price Inertia," mimeo., 1986.

Kuran, Timur, "Asymmetric Price Rigidity and Inflationary Bias," *American Economic Review*, June 1983, *73*, 373–82.

______, "Price Adjustment Costs, Anticipated Inflation, and Output," *Quarterly Journal of Economics*, May 1986, *101*, 407–18.

Mankiw, N. Gregory, "Small Menu Costs and Large Business Cycles: A Macroeconomic Model of Monopoly," *Quarterly Journal of Economics*, May 1985, *100*, 529–37.

Mussa, Michael, "The Welfare Cost of Inflation and the Role of Money as a Unit of Account," *Journal of Money, Credit, and Banking*, May 1977, *9*, 276–86.

______, (1981a) "Sticky Individual Prices and the Dynamics of the General Price Level," in Karl Brunner and Allan H. Meltzer, eds., *Carnegie-Rochester Conference Series on Public Policy: The Costs and Consequences of Inflation*, Autumn 1981, *15*, 261–96.

______, (1981b) "Sticky Prices and Disequilibrium Adjustment in a Rational Model of the Inflationary Process," *American Economic Review*, December 1981, *71*, 1020–27.

Naish, Howard F., "Price Adjustment Costs and Output-Inflation Trade-off," *Economica*, May 1986, *53*, 219–30.

Parkin, Michael, "The Output-Inflation Trade-off When Prices Are Costly to Change," *Journal of Political Economy*, February 1986, *94*, 200–24.

Rotemberg, Julio J., "Aggregate Consequences of Fixed Costs of Price Adjustment," *American Economic Review*, June 1983, *73*, 433–36.

Sheshinski, Eitan and Weiss, Yoram, "Inflation and Costs of Price Adjustment," *Review of Economic Studies*, June 1977, *44*, 287–303.

______ **and** ______, "Demand for Fixed Factors, Inflation and Adjustment Costs," *Review of Economic Studies*, January 1979, *46*, 31–45.

______ **and** ______, "Optimum Pricing Pricing Policy under Stochastic Inflation," *Review of Economic Studies*, July 1983, *50*, 512–29.

Servicing the Public Debt: The Role of Expectations

By GUILLERMO A. CALVO*

We study models in which debt repudiation—openly or through inflation—is possible. The government maximizes the utility of the representative individual, and we focus on no-precommitment equilibria of the Barro-Gordon type. We show cases in which the existence of government bonds generate multiple perfect-foresight equilibria. However, price indexation and/or interest-rate ceilings are shown to be possible solutions of the equilibrium multiplicity problem.

As one looks around the world one is impressed by the wide variety of languages, cultures, and even perceptions of the same phenomena. However, the topic of the government budget is one of the few where the range of issues and degree of emotions appears to find a common denominator. The government budget lies at the basis of political campaigns and plays a significant role in the evaluation of government policy.

Recently, economic theory has begun to catch up with political reality by not only studying the optimality of fiscal and monetary policy in a context where explicit account is taken of the government's budget constraint—the cradle of political disputes—but it has gone a step further by examining the time consistency of optimal policy, that is, the issue of whether it is optimal to keep promises that were optimal to make in the past (for example, during the electoral campaign). The latter lies at the heart of the "credibility" dilemma faced by any serious politician.

Time inconsistency of optimal policy has received a lot of attention in the literature (see, for example, Finn Kydland and Edward Prescott, 1977; Guillermo Calvo, 1978a,b; Robert Barro and David Gordon, 1983a,b; Robert Lucas, Jr., and Nancy Stokey, 1983; Mats Persson, Torsten Persson, and Lars Svensson, 1987, among others). Most of the attention, however, was directed to (a) the identification of the factors responsible for time inconsistency, (b) mechanisms to ensure time consistency, and (c) the characterization of optimal policy when present planners take into account their inability to make future commitments—which leads to the equilibrium concept first examined in macroeconomics by Edmund Phelps and Robert Pollak, 1968). The latter has become the dominant equilibrium concept in this literature (see Kenneth Rogoff, 1987).

Phelps-Polak equilibria were studied and developed much further in several influential articles (for example, Kydland and Prescott, 1977; Barro and Gordon, 1983a,b); most of the examples, however, exhibit unique equilibrium solutions. Non-uniqueness was mentioned in Phelps (1975) and Calvo (1978b), and is one of the central subjects of recent game-theoretic models with an infinite horizon (see Rogoff, 1987), but, in all fairness, it must be said that non-uniqueness tends to be deemed more a theoretical possibility

---

*University of Pennsylvania, Philadelphia, PA 19104-6297. I have greatly benefited from comments received at Johns Hopkins University, Georgetown University, the Macro Lunch Group of the University of Pennsylvania, MIT, and the Research Department of the International Monetary Fund. I am particularly indebted to Matt Canzoneri, Sara Guerschanik-Calvo, Joseph Harrington, Jr., Elhanan Helpman, Nissan Leviatan, Maury Obstfeld, Assaf Razin, Sweder van Wijnbergen, and two anonymous referees. The research on this paper was partly done while I was a visiting scholar at the Research Department of the International Monetary Fund, and it was partially funded by the National Science Foundation. However, any opinions expressed are entirely my own and not those of any of the above institutions.

—reflecting, perhaps, the fact that the models are not entirely correct—rather than a characteristic of empirically relevant models.

The central topic of these notes is the discussion of a set of empirically motivated examples where non-uniqueness of Phelps-Pollak equilibria is clearly exhibited. The possible empirical relevance of the non-uniqueness issue becomes apparent when examining the role of interest-bearing government debt. The events of the 1970s and 1980s suggest very strongly that when governments become strapped for funds, they tend to rely more heavily on bonds' issuance. In fact, Barro (1979) suggests that this may even be optimal if governments are faced with a temporary income fall or a temporary increase in government expenditures. However, the consequent taxation postponement is not free from "credibility" problems: Will the additional debt be paid off in full, will the government find it optimal to resort to higher inflation or currency devaluation to diminish the burden of the debt, etc.? When these questions are posed in terms of more precise language they translate into the following: Is debt repudiation—directly, through interest rate taxes (say), or indirectly through inflation—a characteristic of Phelps-Pollak equilibria? Some answers can be found in Stanley Fischer, 1983; Henning Bohn, 1987; Herschel Grossman and John Van Huyck, 1987; and literature on sovereign countries' debt (see the recent survey by Jonathan Eaton, Mark Gersovitz, and Joseph Stiglitz, 1986). In this paper we pursue the discussion enquiring, specifically, whether the mere existence of government obligations may not be responsible for the existence of multiple Phelps-Pollak equilibria. This appears likely to happen because expected (partial) debt repudiation would tend to be reflected in the interest rate on government bonds (increasing it), while the higher the burden of the debt, the higher would be the temptation to repudiate it. Thus, it should be possible to generate an equilibrium with low interest and low repudiation, coexisting with a high-interest, high-repudiation equilibrium.

Our examples give a strong support to the above conjecture: Multiple solutions are possible even when our examples assume a finite horizon, and the existence of government bonds may be the triggering factor for non-uniqueness.

The implications for policy could be staggering; for, our results suggest that postponing taxes (i.e., falling into debt) may generate the seeds of indeterminacy; it may, in other words, generate a situation in which the effects of policy are at the mercy of people's expectations—gone would be the hopes of leading the economy along an "optimal" path.

For the sake of the exposition, we will first discuss these issues in terms of nonmonetary models with one representative individual. In Section I we will examine a two-period model, where the debt is contracted in the first period and repaid in the second; taxes are distorting and, hence, the government has an incentive to renege on the debt. In order to get an equilibrium with positive public debt, we assume that debt repudiation is costly, and that the cost is proportional to the amount being repudiated. In addition, we assume that individuals know the relevant model; hence, since there is no uncertainty, the equilibrium net rate of interest on public bonds (adjusted for debt repudiation) equals the rate of return on capital (the opportunity cost of private funds). The main result of the section is that if an equilibrium with a positive public debt exists, there are in general two equilibrium points: a "good" Pareto-efficient equilibrium in which there is no debt repudiation, and a "bad" Pareto-inefficient equilibrium where debt is partially repudiated.

In Section II we focus on a monetary economy with non-indexed bonds. Obviously, in this context inflation is equivalent to some form of repudiation, but *negative* repudiation has to be allowed for since price deflation cannot be ruled out. In order to cast the example in terms of a standard monetary model, we follow Barro and Gordon (1983a, b) and modify the previous model by assuming that price changes are costly. This is enough to give results that are qualitatively similar to the ones in the nonmonetary model. Typical of a monetary economy, however, the two equilibria are Pareto inefficient. At the (relatively) good equilibrium, inflation and the nominal interest rate on government bonds are lower than at the bad equilibrium. In addition, we show that if instead the nominal return on government bonds were fully indexed to the price level, then the first-best solution with precommitment can be attained, which suggests that, if inflation is the only means of debt repudiation, indexation of the public debt could lead the economy to a Pareto-superior outcome.

In Section III the analysis is extended to an open economy where foreigners hold a positive share of total public debt. The main purpose of this short section is to establish a link with the relatively well-developed theory of sovereign borrowing. We show that multiple equilibria still can occur in this new context, and that the effect on the bond interest rate of increasing foreigners' participation in that market depends on whether the economy settles down to the good or to the bad equilibrium.

Finally, Section IV summarizes some of the central findings, and discusses some extensions and suggestions for further research.

I. Debt Repudiation

There are two periods: period 0 and period 1, and two types of agents: identical competitive consumers (or individuals) and the government. In period 0 the government borrows b per capita units of output with a (gross) interest factor[1] R_b, that is, in period 1 consumers will receive R_b units of output per unit of bond they hold *and which has not being repudiated*. Therefore, if consumers expect that a proportion θ of total bonds will be repudiated, $0 \leq \theta \leq 1$, the *net* interest factor would be[2]

$$(1-\theta)R_b.$$

We assume consumers can accumulate physical capital with a constant net interest factor equal to $R>1$; hence, in a perfect foresight equilibrium with positive stocks of capital and government bonds, consumers should be indifferent between these two types of assets, and, consequently

$$(1) \qquad (1-\theta)R_b = R.$$

The budget constraint of the government in period 1 is

$$(2) \qquad x=(1-\theta)bR_b+g+\alpha\theta bR_b,$$

where b, x, and g are the per capita stock of bonds, taxes, and government expenditure, respectively, and α stands for the per capita cost per unit of repudiated debt[3] $(0 \leq \alpha < 1)$. Thus, in other words, taxes are required to finance debt repayment, $(1-\theta)bR_b$, government expenditure, g, and the costs of debt repudiation, $\alpha\theta bR_b$. Furthermore, we assume that g is exogenous.[4]

By (2),

$$(3) \qquad \theta bR_b=\frac{bR_b+g-x}{1-\alpha}.$$

In period 1 individuals consume all of their wealth; thus, assuming that government expenditure does not directly affect private consumption in period 1, c, we have

$$(4) \quad c=y-z(x)+kR+(1-\theta)bR_b-x,$$

where y is endowment income, $z(x)$ is a function representing the "deadweight" cost of taxation,[5] and k is per capita physical capital.[6] We assume

---

[1] Let r be an interest rate; we define the corresponding interest factor as $(1+r)$.

[2] Repudiation is a catchall word in this paper that includes anything from open repudiation to a tax on interest.

[3] α could be thought of as transactions costs associated with debt repudiation (legal fees, etc., when repudiation is open). In Section IV other types of costs are discussed. It should be stressed, however, that in this paper we are interested in examining the "mechanics" associated with these types of costs as a prelude to a more substantive analysis where those costs will be subject to a closer scrutiny.

[4] In a more realistic model, the expected net return on capital will also be a function of expected government policy (through expected capital levies, for example). The relationship between the latter and "capital flight" has recently been explored by Eaton (1987).

$$z(0)=z'(0)=0, \tag{5a}$$

$$z''(x)>0 \quad \text{for all } x, \tag{5b}$$

$$\lim_{x\to\infty} z'(x)=\infty=-\lim_{x\to-\infty} z'(x). \tag{5c}$$

The last Inada-type condition will simplify the presentation of our results.

Employing (3) in (4), we get

$$c=y-z(x)+kR+bR_b -(bR_b+g-x)/(1-\alpha)-x. \tag{6}$$

Consequently, a "benevolent" government that tries to maximize c in period 1 subject to its budget constraint, (2), and taking R_b as given (recall that, by assumption, R_b is negotiated at time 0), will choose x so as to minimize

$$z(x)-\frac{\alpha}{1-\alpha}x, \tag{7}$$

subject to (2), and $0\leq\theta\leq 1$. In other words, the problem is to minimize (7) by choosing x such that the associated θ in budget constraint (2) is in the interval $[0,1]$. This condition is equivalent to restricting x in the following manner,

$$g+\alpha bR_b\leq x\leq g+bR_b. \tag{8}$$

Notice that the left-most expression in (8) corresponds to total debt repudiation, $\theta=1$, in which case taxes are equal to government consumption, g, plus the cost of total repudiation, αbR_b. On the other hand, the right-most expression corresponds to no repudiation, $\theta=0$; thus taxes have to be equal to g plus debt repayment.

For the sake of definiteness, we will assume $g>0$ and $b>0$, unless otherwise stated; thus, by (5), if $R_b>0$ (the only relevant case for the model of this section), the minimization of (7) subject to (8) is attained at some unique $x>0$. Notice that, due to (8), the solution will depend on g, and, most importantly for our discussion here, it will also depend on R_b, the interest factor on bonds (which, as argued above, is a predetermined variable in period 1).

The maximum problem for the government in period 1 is depicted in Figure 1.[7] In the figure we portray the two functions associated with constraint (8). Given R_b, the government is thus constrained to choose x between the two straight lines stemming from g on the vertical axis. If x^* is located as in Figure 1, the government will be able to attain the unconstrained maximum x^*, if R_b lies between $\underline{R}$ and $\bar{R}$ (see Figure 1). On the other hand, if $R_b<\underline{R}$, the maximum will be attained on the upper bound (where there is no repudiation, $\theta=0$), while if $R_b>\bar{R}$, the maximum lie on the lower bound (where repudiation is total, $\theta=1$). Consequently, the set of best responses or "reaction function" for the government in period 1, given R_b, is depicted by the heavy line of broken segments in Figure 1. Notice that when $R_b\leq\underline{R}$ the government will be induced to repudiate none of its debt (i.e., set $\theta=0$), whereas if $R\geq\bar{R}$ repudiation will be total (i.e., $\theta=1$). In between $\underline{R}$ and $\bar{R}$, repudiation will be partial, and—by (2), and recalling that in this region $x=x^*$—the repudiation share, θ, will increase with R_b; thus, in other words, *the optimal repudiation share in period 1 is an increasing function of the interest rate contracted in period 0.*

We will be interested in an equilibrium concept in which individuals at time 0 are able to predict the optimal government policy in period 1 (i.e., the optimal government

---

[5] Alternatively, one could follow Barro (1979) and assume $z(x)$ measures tax collection costs; under this interpretation, $z(x)$ should be included in the RHS of (2).

[6] There is no real need to keep track of the nonnegativity of c, because in this model one can always ensure that by setting endowment income, y, sufficiently large.

[7] I am very grateful to Elhanan Helpman for suggesting Figures 1 and 2, which greatly improved the intuitive appeal of the presentation.

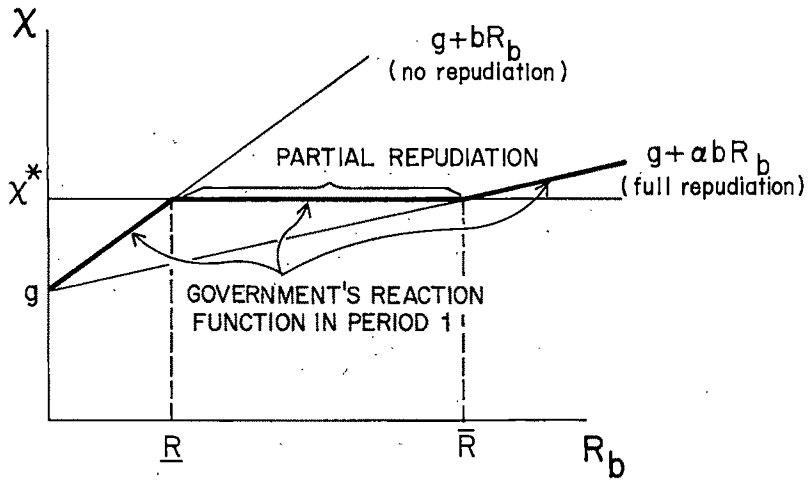


FIGURE 1. DETERMINATION OF GOVERNMENT'S REACTION FUNCTION IN PERIOD 1

decision about x and θ at time 1 as discussed above) with perfect certainty. Under these circumstances (1) holds and, hence, by (2),

$$x = g + (1-\alpha)bR + \alpha bR_b. \tag{9}$$

The latter is a "consistency condition" that must be satisfied by the government's budget constraint at equilibrium, where, by definition, the public can predict exactly the repudiation share, θ, and it is indifferent between holding bonds or capital (i.e., equation (1) is satisfied). Notice, incidentally, that since θ is nonnegative, (1) implies that equation (9) is only relevant over the range where $R_b \geq R$, that is, where the interest rate on public bonds exceeds or equals that of physical capital.

The consistency condition and the government's reaction function are depicted in Figure 2 for the case in which $x^* > g + bR$. Equilibria are found at points of intersection; in the present case these are E^0 and E^1. At E^0 the interest factor on bonds, R_b, is equal to that on capital, R, meaning that the public expects no repudiation; on the other hand, the government's optimal response in period 1 is to set $x = g + bR$, which, as indicated in Figures 1 and 2, lies on the no-repudiation section of its reaction function. Expectations are, thus, fulfilled. Notice that at this equilibrium solution the government would wish to increase x above $g + bR$ and toward x^*, but that is impossible because it would call for setting $\theta < 0$ (negative repudiation), which has been ruled out by assumption.[8]

In the other equilibrium, E^1, repudiation is partial and, thus, the associated interest factor, denoted R_b^1 in Figure 2, is larger than that on capital, R. At this equilibrium, the government is able to attain the unconstrained optimum value of x, x^*.

Our previous discussion covered the case in which $x^* > g + bR$. If, contrariwise, $x^* < g + bR$, the curve depicting the consistency condition would stem from a point like A in Figure 2; it is easy to see that the latter and the government's reaction function will never

---

[8] This type of constraint will be relaxed in Section II. However, one simple way to extend the present example to allow for $\theta < 0$ would be to assume that negative repudiation is also costly. It is easy to see that this would remove any incentive from the government at time 1 to set $\theta < 0$, because the latter would involve bigger tax distortions and repudiation costs than if $\theta = 0$.

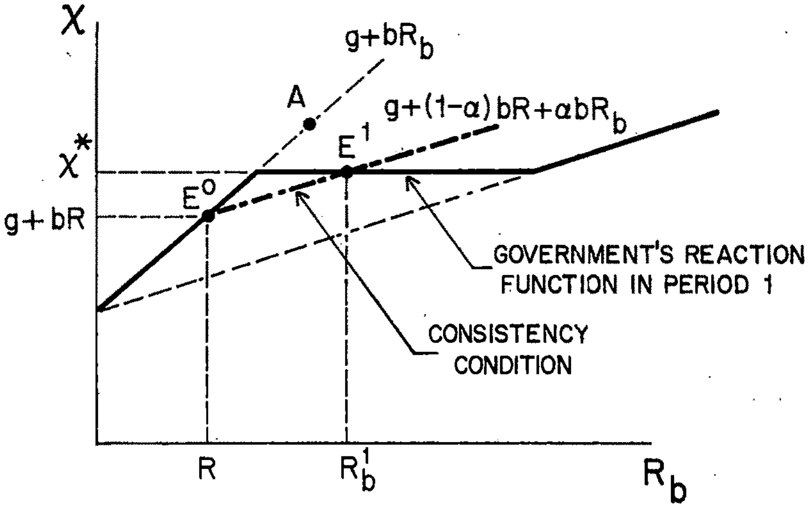

X
$g+bR_b$
A
$g+(1-\alpha)bR+\alpha bR_b$
E^1
X^*
E^0
g+bR
GOVERNMENT'S REACTION FUNCTION IN PERIOD 1
CONSISTENCY CONDITION
R
R_b^1
R_b

FIGURE 2. DETERMINATION OF EQUILIBRIUM

cross, because the slope of the consistency-condition curve is equal to that of the rightmost section of the reaction function. Consequently, there is no equilibrium with a positive stock of bonds.

Finally, in the borderline case where $x^* = g + bR$, equilibrium is unique and $R_b^1 = R = \underline{R}$. The following proposition summarizes our findings:

PROPOSITION 1: *If* $x^* > g + bR$, *then there exist two equilibrium solutions; one of the solutions exhibits no repudiation, while in the other the public debt is partially repudiated. On the other hand, if* $x^* < g + bR$, *there is no equilibrium, and in the borderline case that* $x^* = g + bR$, *equilibrium is unique and there is no repudiation.*

Our model has several interesting implications. In the first place, notice that, by (7), x^* is an increasing function of the cost of repudiation, α; hence, given the initial stock of debt, there is some critical value for the cost of repudiation below which no equilibrium with a positive amount of bonds would exist. The possibility of nonexistence of equilibrium is intuitively clear for the case $\alpha = 0$; for, a benevolent government would repudiate 100 percent of the debt given that there are no direct repudiation costs, and that setting $\theta = 1$ eliminates the deadweight cost of taxes required to service the debt. But, of course, this cannot be an equilibrium with a positive stock of bonds. In addition to confirming this basic intuition, our example shows that the nonexistence problem could arise with a positive α, and that the critical value for the latter depends positively on the value of the outstanding debt.

The most important implication of the analysis, however, is that problems do not necessarily go away when repudiation costs are set above the critical level; for, in such a case the economy will normally exhibit two equilibria: a nonrepudiation equilibrium in which taxes equal $g + bR$ (recall Figure 2), and another equilibrium in which the debt is partially repudiated and taxes are set at $x^* > g + bR$. Notice that, by (1) and (4), in equilibrium we have

$$c = y - z(x) + (k + b)R - x. \tag{10}$$

Therefore, these two solutions can be Pareto-ranked, and the nonrepudiation solution is the dominant one. Unfortunately, however, without further restrictions the economy could end up at any one of these equilibria. The interest factor in the "good" equilibrium is always R; however, the value of R_b and θ at the "bad" equilibrium (E^1 in Figure 2) are sensitive to changes in b. Somewhat paradoxically, as it can easily be established employing Figure 2, at the bad equilibrium *the bond interest factor is smaller the larger is the total outstanding government debt*. It should be noted, however, that the consumption cost of being at the bad equilibrium is independent of total debt[9] (because $x = x^*$, and the latter does not depend on b). But, on the other hand, x^* depends positively on α; therefore, our analysis reveals that while repudiation costs are important for ensuring the existence of equilibria with a positive stock of bonds, by (5), (7), and (10), the larger are the repudiation costs, the larger will also be the consumption cost associated with the bad equilibrium. Thus, *although increasing the costs of repudiation enhances the "credibility" of the government in the capital markets, the latter may be bought at the cost of lower consumption if the economy ends up at the bad equilibrium.*

The way the above "game" has been structured is consistent with situations in which the government at time 0 auctions off b, and lets the market free to determine the associated interest factor, R_b. An alternative[10] would be for the government at time 0 to refuse to sell bonds at interest factors which are equal to or exceed R_b^1 (recall Figure 2). Thus, under our assumptions it follows that the only possible equilibrium would be $R_b = R$, the first best. This is an important observation which suggests that free auctioning of the public debt may be dominated by a system in which the government stops selling bonds after the implied interest rates go beyond certain bounds. This observation will take up further relevance when we discuss it in the context of the next section.

II. Money and Non-Indexed Debt

The concept of partial repudiation takes a more familiar form in the context of a monetary model, because when variables are expressed in real terms, changes in the rate of inflation imply changes in the real value of assets which are not indexed to the price level. Two important assets of this kind are (high-powered) money and nominal government debt. The literature has paid a great deal of attention to the former (for example, Barro and Gordon, 1983a, b); Lucas and Stokey, 1983; Calvo (1978a, b), but there is growing awareness about the importance of understanding the role of nominal debt (for example, Henning Bohn 1987; Grossman and Van Huyck, 1987; Persson, Persson, and Svensson, 1987). As far as I know, however, none of the available papers addresses the issue of non-uniqueness of equilibria, which is the main focus of the following discussion.

To minimize the use of new notation, we redefine R_b as the nominal interest factor on nominal non-indexed bonds (i.e., $R_b = 1 + i$, where i is the nominal interest rate from periods 0 to 1). Thus, if we let P_t stand for the price level in period $t = 0, 1$, then the *real* interest factor on period-0 bonds would be

$$R_b P_0 / P_1. \tag{11}$$

We can, therefore, think of the ratio P_0/P_1 as the share of the debt which is not repudiated in period 1; hence, using the notation of previous section, we will write

$$P_0/P_1 = (1-\theta). \tag{12}$$

We define the rate of inflation between periods 0 and 1, π, as follows,

$$\pi = \frac{P_1 - P_0}{P_0}. \tag{13}$$

Thus, by (12) and (13)

$$\theta = \frac{\pi}{1+\pi}. \tag{14}$$

---

[9] This invariance does not hold in the model discussed in Section IV.

[10] This fine point was suggested to me by Assaf Razin.

Consequently, each rate of inflation, π, $-1 < \pi$, is associated with a unique rate of repudiation, θ. We are thus entitled to carry out our discussion in terms of θ instead of π. Notice, however, that contrary to the previous section, θ could be negative because it is only constrained to satisfy

$$(15) \qquad -\infty < \theta \leq 1.$$

In line with our previous results, it is easy to convince oneself that if the costs of inflation/repudiation are nil, then a social welfare maximizer will repudiate 100 percent of the debt, that is, he will set $\theta = 1$—or, equivalently, $\pi = \infty$. Therefore, the existence of a well-defined interior equilibrium solution requires, once again, the introduction of (social) repudiation costs. As a matter of fact, one could translate the model of Section II in monetary terms using (14), and interpret those costs as being related to, for example, the Olivera-Tanzi effect. However, I believe it will be more useful to develop the analysis in terms of a more conventional monetary model in which inflation costs directly affect the consumers' utility functions (see Barro and Gordon, 1983a,b), and where explicit account is taken of the inflation tax on high-powered money.

Let the supply of high-powered money in period t, $t = 0,1$, be denoted by M_t. For the sake of simplicity, we assume that the monetary authorities can directly determine the price level P, by, for example, setting the exchange rate. Furthermore, we assume that the demand for money satisfies[11]

$$(16) \qquad M/P = \kappa, \qquad \kappa > 0$$

The revenue from inflation is conventionally defined as the amount of real resources that the government can obtain by the associated sale of high-powered money; thus, in period 1 it would be given by

$$(17) \qquad \frac{M_1 - M_0}{P_1} = (M_1/P_1) - (M_0/P_0)(P_0/P_1) = \kappa\theta,$$

where the last equality follows from (12) and (16).

The government budget constraint in period 1 is given by

$$(18) \qquad x = (1-\theta)bR_b + g - \kappa\theta.$$

This is similar to (2) above, except that inflation is assumed not to be costly for the government, and the inflation tax is subtracted from total government expenditure (inclusive of debt service) in order to calculate the amount of required conventional taxes.

Effective consumption, c, is defined as follows,

$$(19) \qquad c = y - z(x) + kR + (1-\theta)bR_b - x - \kappa\theta - \mathfrak{R}(\theta),$$

where $\mathfrak{R}(\cdot)$ is the inflation-cost function. Again, this is very similar to (4) above, except that we have quite naturally subtracted the inflation tax, and, as discussed at the outset, we assume that inflation is reflected in a direct welfare cost for consumers.

We assume

$$(20a) \qquad \mathfrak{R}(0) = \mathfrak{R}'(0) = 0,$$

$$(20b) \qquad \mathfrak{R}''(\theta) > 0 \quad \text{for all } \theta.$$

We will first study the optimal problem with precommitment, that is, the problem of maximizing c with respect to θ *in period* 0. Since we maintain the assumption of perfect foresight, equation (1)—the Fisher equation—holds. Therefore, the above-mentioned problem can be formally stated as follows:

$$\max_{\theta \leq 1} c,$$

---

[11] The following simple form is sufficient for conveying the central insights of this section. Extensions are possible; however, see fn. 14.

subject to (1), (18), and (19).[12] The latter is equivalent to

$$\text{(21)}\quad \underset{\theta\le 1}{\text{Min}}\left[z(g+bR-\kappa\theta)+\Re(\theta)\right],$$

which, by (5) and (20) has a unique solution, denoted θ^{FB} (for "first best"). Since the function in (21) is strictly convex, one can readily prove the following useful proposition:

PROPOSITION 2: *Assume that* (1) *holds, and let* c^i *be the effective consumption associated with* θ^i, $i=1,2$; *then, if* $\theta^{\text{FB}}\le\theta^1<\theta^2$, *we have that* $c^2<c^1$.

In other words, the above proposition states that if inflation is higher than its first-best level, then social welfare—as measured by effective consumption— is a decreasing function of inflation.

We now turn to the second-best situation which is the focus of our analysis. The government is not able to precommit the inflation level, and, therefore, the government maximizes social welfare in period 1, taking the interest rate factor R_b and period-0 inflationary expectations as given. Recalling (19), this problem is equivalent to

$$\text{(22)}\quad \underset{\theta\le 1}{\text{Min}}\,F(\theta;R_b)$$

$$\equiv \underset{\theta\le 1}{\text{Min}}\left[z\left(g+b(1-\theta)R_b-\kappa\theta\right)+\Re(\theta)\right].$$

Notice that in the first-best problem (21) the government takes (1) into account because it is maximizing from the perspective of period 0 and consumers would not be in an (interior) equilibrium if the Fisher equation does not hold. In (22), however, the government controls θ taking R_b as given, because it maximizes effective consumption in period 1 in an environment in which R_b has been negotiated in the previous period (i.e., period 0), and, by assumption, R_b is not contingent on future policy or events.

By (22), at an interior[13] (second-best) optimum we have

$$\text{(23)}\quad F_\theta\equiv-z'(x)(bR_b+\kappa)+\Re'(\theta)=0,$$

implying, by (5) and (20) that optimal $\theta>0$. The second-order condition is always satisfied because F is strictly convex with respect to θ. Also, notice that, by (5) and (20), the minimization problem (22) always has a solution, and by strict convexity, the solution is unique; we denote it θ^{SB} (for "second best"). By (23), there exists some function $\varphi(\cdot,\cdot)$, such that

$$\text{(24)}\qquad \theta^{\text{SB}}=\varphi\left(\overset{+}{bR_b},\overset{+}{g}\right),$$

where signs over the variables indicate those of the corresponding partial derivatives at an interior solution. Consequently, the (second-best) rate of inflation is an increasing function of nominal debt service, bR_b, and government expenditure, g.

An important implication of the above is that if there exists a positive stock of government bonds, b, then θ^{SB} is an increasing function of the nominal interest rate factor, R_b; but, by Fisher equation (1), equilibrium θ is also an increasing function of R_b and thus—like in Figure 1—the two schedules (i.e., the government's reaction function φ and the Fisher equilibrium relationship) are upward sloping, which, once again, opens the door for multiple equilibria.

It is important to notice that in the absence of government debt (i.e., $b=0$), equilibrium will be unique because θ^{SB} would be independent of R_b. Thus, in the present simple context, the familiar inflation tax on high-powered money plays no significant role in causing non-uniqueness.[14]

---

[12] This is like the optimization problem examined in Phelps (1973). Its solution differs from Milton Friedman's Optimum Quantity of Money ($\theta=0$ in the present context) because, contrary to Friedman, we assume taxes are distorting.

[13] In what follows we will concentrate on interior solutions (i.e., $\theta<1$), corresponding to cases in which the inflation rate is less than infinity (recall (14)).

[14] This result could be extended to the case in which the demand for money is sensitive to the rate of interest

In order to complete the non-uniqueness argument, the next step will be to show a simple example of non-uniqueness when $b > 0$. Consider the case in which

$$\Re(\theta) = \frac{\beta}{2}\theta^2. \tag{25}$$

Therefore, by (1) and (23), we have

$$z'(x) = \beta \frac{\theta}{bR/(1-\theta)+\kappa}, \tag{26}$$

and, by (1) and (18), we get

$$x = bR + g - \kappa\theta. \tag{27}$$

Equilibrium relationships (26) and (27) are depicted in Figure 3 for the limit case in which the demand for money is zero, that is, $\kappa = 0$. By continuity, it is quite clear that multiplicity of solutions would continue to hold for κ in some neighborhood of 0. It is, however, very interesting to note that in the present example uniqueness would prevail if κ is sufficiently large.[15] This is perfectly consistent with our previous observation that the existence of high-powered money is conducive to uniqueness. The result is also highly suggestive; since the value of κ may have a lot to do with the inflationary history of a country, the above result indicates that *multiple solutions may be a bigger problem for countries which have recently suffered from high inflation*.[16]

By (21), the first-order condition for a first-best optimum is

$$-z'(g + bR - \kappa\theta)\kappa + \Re'(\theta) = 0. \tag{28}$$

Thus, recalling that in Section II we assumed $g > 0$ and $b > 0$, we have $\theta^{\mathrm{FB}} > 0$. On the other hand, recalling (18), the first-order condition for a second-best optimum is at equilibrium—that is, when (1) holds—

$$-z'(g + bR - \kappa\theta)(bR + \kappa) + \Re'(\theta) = 0. \tag{29}$$

Notice that, by (5), (20), (28), and (29), the first- and second-best optima satisfy $\kappa\theta < g + bR$, implying that in neither one of these optima will the inflation tax be enough to finance total government expenditures. Consequently, by (5), (20), (28), and (29), we have, at equilibrium,

$$\theta^{SB} > \theta^{FB}. \tag{30}$$

In other words, we have shown that optimal inflation with precommitment is lower than in any of the equilibrium solutions obtained under no precommitment. Consequently, recalling Proposition 2, we deduce that equilibria without precommitment can be ranked according to their associated inflation rates: given any two of these equilibria, the one exhibiting the largest rate of inflation yields the smallest social welfare (as measured by effective consumption, see (19)). In terms of the example associated with Figure 3, therefore, we are entitled to call θ^0 the "good," and θ^1 the "bad" equilibrium.[17]

Price indexation of government debt has important effects on this economy since it removes the inflation incentives associated with the debt burden. With full-price indexa-

---

if the elasticity of $z'(x)$ with respect to x is less than unity, and the inflation tax does not exceed regular taxes, the normal case. Our proof, however, relies on the existence of more than two periods, and will, therefore, not be presented here. For an infinite-horizon model of non-uniqueness along the present lines, but without government bonds, see Maurice Obsfeld (1988).

[15] The exact condition is $\kappa > bR + g$, that is, real monetary balances must exceed total government expenditure, including the inflation-adjusted service on the public debt.

[16] This point emerged during a useful conversation with Sweder van Wijnbergen.

[17] Multiple equilibria is by no means a necessary feature of these models. Uniqueness would, in fact, prevail if, for example, $\Re(\theta) = \pi^2$; the latter is related to the work of Barro and Gordon (1983a,b) and Bohn (1987). For a somewhat detailed discussion of these issues, see Calvo (1987).

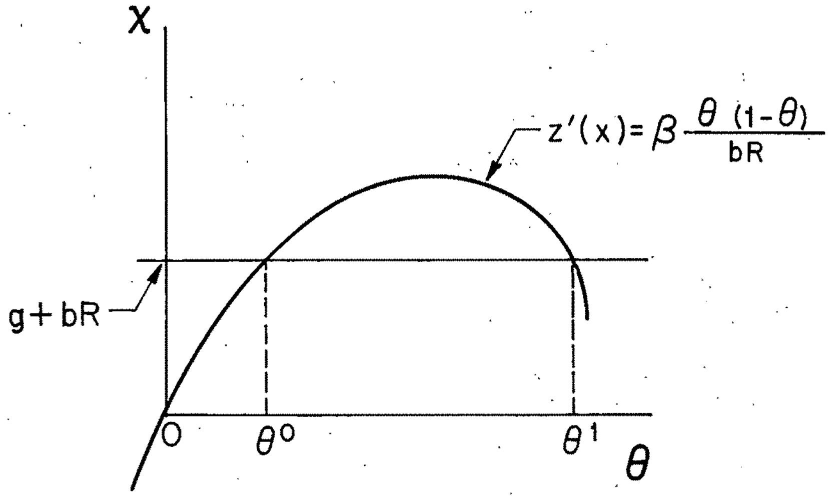


FIGURE 3. MONETARY EXAMPLE

tion and an interest-insensitive demand for money (i.e., constant κ, as in the present model), it is easy to see that the no precommitment equilibrium coincides with the first best.[18]

An alternative to indexation—which was discussed at the end of Section I—is to put bounds on interest rates; this could ensure that the solution settles at θ^0 in Figure 3 rather than at θ^1. However, one can easily show that such a solution is inferior to bond indexation, because the latter results in lower inflation[19] (recall Proposition 2 and (30)).[20]

As the example depicted in Figure 3 shows, an increase in the cost of inflation (i.e., an increase in β, recall (25)), has ambiguous results on welfare. A rise in β causes the curve in Figure 3 to shift upward, raising inflation in the bad equilibrium, and lowering it in the good one. Clearly, net consumption, c, will decrease in the bad equilibrium (because β and inflation go up). In the good equilibrium (i.e., θ^0), on the other hand, the picture is less clear, because inflation goes down, which tends to be welfare improving. However, if $\kappa = 0$, a simple manipulation using (26) and (27) shows that welfare increases with β. Thus, as in Section I, the effects of increasing the costs of inflation/repudiation are ambiguous: they could be welfare-improving in the good equilibrium, but they definitely lower welfare in the bad equilibrium.

Our example sheds some light on the issue studied by Bennett McCallum (1984) of whether monetary policy could be separated from fiscal considerations. His discussion has to do with the long-run (literally when $t \to \infty$) feasibility of policies that keep money supply at a constant level while running a fiscal deficit. In our model the relationship between the rate of inflation and fiscal policy is very direct for the case of non-indexed bonds (still the dominant kind of government bonds in the United States and several other countries), and confirms and complements McCallum's insight that there are no grounds for expecting fiscal policy to have no monetary implications even though, for instance, the monetary authority promises to keep a constant price level (or a constant money supply in a setup where the latter is explicitly modeled). In our model this happens because the government may not commit itself *credibly* to zero inflation in the presence of non-indexed debt, since the level of the latter determines the set of equilibrium inflation rates that could be sustained *ex post*.[21] This relationship between the stock of government bonds and inflation, incidentally, will hold true even when the equilibrium is unique (see Bohn, 1987).

---

[18] Indexation would not suffice for attaining the first-best solution if the demand for money is sensitive to the rate of interest. I conjecture, however, that there are relevant examples where one could still show that debt indexation increases social welfare.

[19] This is not intended to be a thorough analysis of the welfare economics of bond indexation since there are some important aspects of the issue which have not been taken into account. For example, as pointed out to me by Nissan Liviatan, it has been argued that indexation may reduce the base of the inflation tax because in such a case bonds become better substitutes for domestic money. This aspect is not covered by our discussion because κ was assumed to be independent of the degree of indexation.

[20] Bounds on interest rates may be particularly difficult to implement when there exists an active private capital market. See Calvo (1987) for some discussion on this issue.

Our results add a pessimistic note to the view that price controls might help leading the economy to the "good" solution,[22] which appears to have played some role in the design of the recent "heterodox" stabilization plans in Argentina, Brazil, and Israel. This is so, because in our example the government has direct control on the price level, but, once again, non-uniqueness arises due to the fact that its *ex post* behavior is not immune to *ex ante* expectations. As a matter of fact, our previous discussion suggests that interest-rate pegging of some sort may be as much a crucial ingredient of a credible stabilization effort as price controls are, and that interest rate pegging could in some occasions be all that is needed for credibility.

III. International Debt. A Short Detour

Borrowing and lending among sovereign states has recently received considerable attention (see, for instance, the recent survey by Eaton, Gersovitz, and Stiglitz, 1986), so a few words establishing the connection between the present paper and that type of literature may be in order.

A common characteristic of our examples is that the "size of the penalty" (i.e., the cost of repudiation) is a function of the degree of repudiation; in the example of Section II the cost was proportional to the value of the debt, whereas in that of Section III the cost was simply a function of the fraction that was repudiated. In the international debt literature, on the other hand, it is typically assumed that the cost of repudiation bears no relationship to the size of the debt. This explains why it was relatively so easy for us to generate multiple equilibria while in that literature equilibrium is typically unique.[23]

The presence of foreign debt adds an interesting twist to our previous models. Thus, for example, we could assume that only a proportion γ, $0 \leq \gamma \leq 1$, of the total public debt is owned by domestic residents, while $(1-\gamma)$ is owned by foreigners whose welfare is of no concern to the local authorities. Thus equation (4) now becomes

$$c = y - z(x) + kR + \gamma(1-\theta)bR_b - x, \tag{4'}$$

and (19) is similarly modified. One can now examine the relationship between γ and the equilibrium outcomes. This is a straightforward exercise which, however, shows that the results depend very strongly on the model. In the model of Section I, an increase in foreign ownership of the public debt (i.e., a fall in γ) induces a fall in the bond interest rate at the bad equilibrium, but no change at the good one. On the other hand, in the monetary model of Section II (with $\kappa = 0$), interest rates in both equilibria are affected, rising in the bad equilibrium and falling in the good one. There does not seem, therefore, to be a clear-cut relationship between the participation of foreigners in the official bonds market, and the degree of repudiation or inflation.

---

[21] Notice that our arguments establish a link between the stock of non-indexed debt and inflation, in contrast with McCallum (1984) who emphasized the relation between fiscal deficits and inflation.

[22] For recent models that could be employed to support that point of view, see Michael Bruno and Fischer (1985) and Miguel Kiguel (1986).

[23] An exception is the recent work of Eaton (1987), where, however, the issue of debt repudiation is not central to the analysis.

IV. Concluding Remarks

The central message that comes across this paper is that expectations may play a crucial role in the determination of equilibrium when the government debt is auctioned off to the public, and there is no attempt to manage expectations or to peg interest rates on the government debt. In our examples, we saw that the nominal interest rate is not simply a passive reflection of people's inflationary expectations, but rather that the nominal interest rate is actually one of the main determinants of inflation. Consequently, a credible anti-inflationary policy would have to implement rules to prevent nominal interest rates to become unduly high. The two simple mechanisms that were suggested in our discussion were: (a) price indexation of the public debt, and (b) refraining from issuing new government bonds when their interest rate exceeds some well-defined bounds. In more general terms, however, our discussion pointed out to the advisability of governments taking a more active stance on nominal interest rates. In my opinion, this implication of the formal analysis is likely to be particularly relevant in situations where the stabilization program is launched in the midst of very high inflation.

The analysis has been conducted in terms of models with one representative consumer because so much of the related literature has been couched in this form (prominently Barro, 1979; and Lucas and Stokey 1983). However, I suspect that one of the important reasons for debt repudiation to be costly (other than due to "reputation" costs) is the fact that not all individuals are alike and, therefore, the incidence of repudiation is not uniform.[24] Consider, for example, a model of the type discussed by Thomas Sargent and Neil Wallace (1982) where only a subset of total population holds non-interest bearing public debt (cash, say), while the rest of society holds high-denomination non-indexed bonds. For the sake of the argument, suppose, in addition, that everybody pays the same amount of taxes. Clearly, if the planner adheres to utilitarian principles, for example, debt repudiation could be costly (or beneficial), because the latter would be an instrument of wealth redistribution. In that context, the planner may end up repudiating less than 100 percent of the debt (held by the "rich," say) because of the loss it would cause cash-holders (the "poor"). I think the analysis of these types of models should provide us with a deeper understanding of these issues. It is my feeling, however, that the central insights of this paper will carry over to those models.

Another interesting extension would be to introduce uncertainty. Notice, first, that the above two-period examples would not exhibit random equilibria, because once the interest rate is determined in period 0, there exists only one optimum response for the government in period 1. An easy extension, however, would be to make the repudiation cost, α, random. Thus, if individuals are risk neutral, for example, R_b *times* the expected repayment share would in equilibrium equal the interest factor on capital R; the repudiation shares for each state of nature, on the other hand, would depend on R_b (as in our previous analysis). Therefore, the equilibrium values of R_b will be deeply intertwined with the probability distribution of α. Clearly, since R_b is just one number, we will observe situations where *ex post* the *actual* net return on bonds exceeds (or falls short of) that of physical capital. Consequently, these models could help explain phenomena like the high *ex post* real interest rates observed at the beginning of anti-inflationary programs (see, for instance, Rudiger Dornbusch, 1985, and Jeffrey Sachs, 1986).

Finally, an important extension of the analysis would be to allow for more than two periods,[25] and the simultaneous existence of explicit repudiation (for example, interest rate taxes) and inflation.

---

[24] See Carol Rogers (1986) and Calvo and Obstfeld (1988) for models which highlight the link between time inconsistency and wealth distribution.

[25] For some progress on this front, see Calvo (1987).

REFERENCES

Barro, Robert J., "On the Determination of the Public Debt," *Journal of Political Economy*, October 1979, *87*, 940–71.

______, **and Gordon, David B.**, (1983a) "A Positive Theory of Monetary Policy in a Natural Rate Model," *Journal of Political Economy*, August 1983, *91*, 589–610.

______ **and** ______, (1983b) "Rules, Discretion and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, July 1983, *12*, 101–21.

Bohn, Henning, "Why Do We Have Nominal Government Debt?," Department of Finance, The Wharton School, University of Pennsylvania, Revised Version, April 1987, *Journal of Monetary Economics*, forthcoming January 1988, 127–40.

Bruno, Michael and Fischer, Stanley, "Expectations and the High Inflation Trap," unpublished manuscript, September 1985.

Calvo, Guillermo A., (1978a) "Optimal Seigniorage from Money Creation: An Analysis in Terms of the Optimum Balance of Payments of Problem," *Journal of Monetary Economics*, August 1978, *4*, 503–17.

______, (1978b) "On the Time Consistency of Optimal Policy in a Monetary Economy," *Econometrica*, November 1978, *46*, 1411–28.

______, "Staggered Prices in a Utility Maximizing Framework," *Journal of Monetary Economics*, September 1983, *12*, 383–98.

______, "Controlling Inflation: The Problem of Non-Indexed Debt," *Debt Adjustment and Recovery: Latin America's Prospect for Growth and Development*, in S. Edwards and F. Larraín, eds., New York: Basil Blackwell, forthcoming.

______, **and Obstfeld, Maurice**, "Optimal Time Consistent Fiscal Policy with Finite Lifetimes: Analysis and Extensions," CARESS Working Paper No. 87-09, University of Pennsylvania, March 1987. A shorter version with the same main title was published in *Econometrica*, March 1988, *56*, 411–32.

Dornbusch, Rudiger, "Stopping Hyperinflation: Lessons from the German Experience in the 1920's," NBER Working Paper No. 1675, May 1985.

Eaton, Jonathan, "Public Debt Guarantees and Private Capital Flight," *The World Bank Economic Review*, May 1987, *1*, 377–96.

______, **Gersovitz, Mark and Stiglitz, Joseph E.**, "The Pure Theory of Country Risk," *European Economic Review*, June 1986, *30*, 481–513.

Fischer, Stanley, "Welfare Aspects of Government Issue of Indexed Bonds," in R. Dornbusch and M. H. Simonsen, eds., *Inflation, Debt, and Indexation*, Cambridge, MA: MIT Press, 1983, 223–46.

Friedman, Milton, *The Optimum Quantity of Money and Other Essays*, Chicago: Aldine Publishing, 1969.

Grossman, Herschel I. and Van Huyck, John B., "Nominally Denominated Sovereign Debt, Risk Shifting, and Reputation," NBER Working Paper Series No. 2259, May 1987.

Kiguel, Miguel, "Stability, Budget Deficits, and the Monetary Dynamics of Hyperinflation," unpublished manuscript, June 1986.

Kydland, Finn and Prescott, Edward C., "Rules Rather Than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, June 1977, *85*, 473–91.

Lucas, Robert E., Jr. and Stokey, Nancy L., "Optimal Fiscal and Monetary Policy in an Economy Without Capital," *Journal of Monetary Economics*, July 1983, *12*, 55–94.

McCallum, Bennett T., "Are Bond-Financed Deficits Inflationary? A Ricardian Analysis," *Journal of Political Economy*, February 1984, *92*, 123–35.

Obstfeld, Maurice, "A Theory of Capital Flight and Currency Depreciation," unpublished manuscript, January 1988.

Persson, Mats, Persson, Torsten and Svensson, Lars E. O., "Time Consistency of Fiscal and Monetary Policy," *Econometrica*, November 1987, *55*, 1419–32.

Phelps, Edmund, S., "Inflation in the Theory of Public Finance," *Swedish Journal of Economics*, March 1973, *75*, 67–82.

______, "The Indeterminacy of Game-Theoretic Equilibrium in the Absence of an Ethic," in *Altruism, Morality and Economic Theory*, E. S. Phelps, ed., New York: Russell Sage Foundation, 1975, 87–106.

______ **and Pollak, Robert A.**, "On Second-Best

National Saving and Game-Equilibrium Growth," *Review of Economic Studies*, April 1968, *35*, 185–99.

Rogers, Carol Ann, "The Effect of Distributive Goals on the Time Inconsistency of Optimal Taxes," *Journal of Monetary Economics*, 1986, *17*, 251–69.

Rogoff, Kenenth, "Reputation, Coordination and Monetary Policy," The Hoover Institution, Working Paper in Economics E-87-14, March 1987.

Sachs, Jeffrey, "The Bolivian Hyperinflation and Stabilization," NBER Working Paper No. 2073, November 1986.

Sargent, Thomas J. and Wallace, Neil, "The Real-Bills Doctrine versus the Quantity Theory: A Reconsideration," *Journal of Political Economy*, December 1982, *90*, 1212–36.

Social Contracts as Assets: A Possible Solution to the Time-Consistency Problem

By LAURENCE J. KOTLIKOFF, TORSTEN PERSSON, AND LARS E. O. SVENSSON*

We present a new solution to the time-consistency problem that appears capable of enforcing ex ante policy in settings where other enforcement mechanisms do not work. The solution involves a social contract that specifies the optimal ex ante policy and is effectively sold by successive old generations to successive young generations, who pay for the social contract through intergenerational transfers. Both old and young generations have an economic incentive to fulfill the social contract.

The requirement that policies be *ex post* optimal, or time consistent, in general leads to outcomes that are suboptimal relative to those that binding commitments would sustain. A classical example, considered by Stanley Fischer (1980), is capital taxation in a two-period model. A second-period capital levy is first best from the perspective of period two, but third best from the perspective of period one. If there is nothing to enforce a government commitment against second-period capital taxation, the *ex post* desire to engage in first-best taxation drives the economy from a second-best to a third-best outcome.

Researchers have suggested a variety of ways to enforce commitment to *ex ante* optimal policies. Finn Kydland and Edward Prescott (1977) argue for rules rather than discretion. A number of writers—see Kenneth Rogoff (1987) for a survey—introduce government reputation as a mechanism to enforce commitment. Robert Lucas and Nancy Stokey (1983), and Mats Persson, Torsten Persson, and Lars Svensson (1987) show that, assuming partial commitment to honoring its bonds, governments can deter deviations from *ex ante* policies by devising financial instruments whose market values fall if *ex post* policy deviates from *ex ante* policy.[1]

This paper presents a solution to the time-consistency problem that appears capable of enforcing *ex ante* policy in a variety of settings in which other enforcement mechanisms do not work. The solution involves formulating a "social contract" (or institution or agreement) that specifies the *ex ante* optimal policy and that can be "sold" by successive old generations to successive young generations. Each young generation pays for the social contract by paying a larger share of taxes than it would otherwise do. Both old and young generations have an economic incentive to fulfill the social contract. For the old generation that owns the social contract, breaking the social contract makes it valueless, and the generation suffers a capital loss. For the young generation the

---

*Kotlikoff: National Bureau of Economic Research, 1050 Massachusetts Ave., Cambridge, MA 02138, USA. Persson and Svensson: Institute for International Economics, University of Stockholm, S-106 91 Stockholm, Sweden. Persson and Svensson are grateful for support from National Science Foundation grant no. SES-8605871. Svensson thanks NBER for its hospitality during his visit there in July and August 1986. We appreciate helpful comments by Gary Becker, Michael Cadwalader, Oded Galor, Louis Kaplow, Robert Lucas, Bentley McLeod, John Moore, Charles Wilson, and seminar participants at Brown University, Columbia University, Duke University, IIES, NBER, New York University, Princeton University, Queens University, Tel-Aviv University, University of Chicago, University of Pennsylvania, University of Rochester, and Yale University. We thank Kerstin Blomquist and Maria Gil for secretarial and editorial assistance.

[1] Persson (1988) gives a broad survey of recent research on time-consistency (credibility) problems.

economic advantage of purchasing the social contract exceeds its price as well as the economic gain from setting up a new social contract.

Fischer's model extended to include overlapping generations provides a convenient device for illustrating the potential asset nature of social contracts. Assume, to keep matters simple, that only the elderly consume the public good. Since the old and young have no common interests they live side by side, but have separate governments called councils. These councils are democratically elected each period, and since the members of any generation are identical, each council simply carries out the current unanimous wishes of its constituents. Each generation's council has the right to tax its constituents and is obliged to supply the public good when its constituents are old. In fact our model makes each generation completely economically independent. This feature of the model obviously abstracts from important aspects of actual policymaking problems. But we wish to focus here on the rôle of social contract in enforcing desirable policies from an efficiency point of view. This is best done by abstracting from redistribution directly by taxation of other generations, and indirectly by policy-induced changes in factor prices.

Assume now that the *ex ante* second-best, but time-inconsistent tax structure involves relatively low capital taxation. In contrast, the third-best, time-consistent tax structure clearly involves higher capital taxation, because each individual when old will wish his council to levy an *ex post* nondistortionary tax on capital. Since the capital tax will be anticipated when young, saving will be distorted, implying a third-best rather than second-best level of welfare.

In this model each generation is self-governing and cannot be bound by any "rules." Reputation in the context of incomplete information about the policymaker—as in David Backus and John Driffill (1985)—requires, as far as we can see, the policymaker to be an agent separate from its constituency. In our model we consider the government simply as a representative of its constituency, with the same preferences as the identical consumers it represents. Reputation, in the incomplete-information interpretation therefore does not work. Reputation in the context of complete information about the policymakers—as in Robert Barro and David Gordon (1983)—requires the policymakers, planning horizon to be effectively infinite. Thus, it does not work here since policymakers solely represent their generation who has a finite horizon. Nevertheless, our solution resembles the (complete information) reputational mechanism in some ways. Finally, the partial commitment approach, while also resembling our solution in some respects, requires invoking a certain degree of precommitment. Since such precommitment is in no way suggested by the model, it is not a very satisfying solution.

Enforcement through the sale of a social contract can arise in the following way. Suppose one particularly enterprising leader of the young, after tedious discussions with her electorate, drafts a social contract which states that the generation holding the social contract is exempt from capital taxation (above a certain level). The enterprising leader of the young ensures her electorate that in the next period, when their generation (called the first generation) is old, she will approach the council of the young (called the second generation) and offer to sell the social contract in exchange for tax revenue from the young to help pay for the public good of the old. She argues convincingly that after agreeing on an acceptable amount of tax revenue, the council of the second-generation young will purchase the social contract. The reason is that (1) the social contract will protect the second generation from excessive (and third best) capital taxation because selling the social contract to the subsequent young (the third generation) will be more beneficial in terms of additional tax receipts than the second-period efficiency gains from capital taxation, and (2) the cost of purchasing the social contract less its resale value is less than the transactions' costs to the second generation of negotiating about the creation of their own social contract.

As demonstrated below, there is, in fact, an equilibrium in which the social contract is passed on from one generation to the next, no generation wishes to abrogate the social contract, and, if the social contract is destroyed by chance; it will immediately be reconstituted. The result is that an outcome with low capital taxes and high saving, almost as good as the second-best outcome, can be enforced.

In contrast to the analyses of Lucas and Stokey (1983), and Persson, Persson, and Svensson (1987) in which potential capital losses on government financial assets deter deviations from *ex ante* policy, the assets in this paper are agreements which may or may not be written down explicitly. Thus, in our example, no physical contract or certificate is required, simply an understanding that the quid pro quo for the old to maintain the tradition (institution) of no capital taxation is for the young to contribute the appropriate share to pay for the public good of the old. Probably the paper's closest antecedent is Jonathan Eaton's (1985) analysis of banks' repayment enforcement. In Eaton's paper failure of banks to pursue defaulters leads to a reduction in their stock values.[2]

The paper examines the sale of social contracts in a simple linear model that yields analytical expressions for second- and third-best tax structures. The linear model is attractive because excess burdens are additive; indirect utility in distorted regimes differs from the first-best level of utility by the sum of excess burdens. Section I presents the model, while Section II presents the unobtainable first-best optimum, the unenforceable second-best optimum, and the time-consistent third-best optimum. Section III demonstrates how a "salable" social contract comes into existence and how a social contract can enforce an outcome close to the second best. These points are illustrated with a numerical example. Section IV deals with enforcement without transaction costs. Section V discusses extensions and broader implications of our results.

I. The Model

The model is a life-cycle overlapping generations model with production and investment. There is one good and labor. Production is linear, and output of goods in period t, Y_t, depends on the input of capital in period $t-1$, K_{t-1}, and labor in period t, L_t, according to

$$(1)\quad Y_t = RK_{t-1} + wL_t, \quad R > 1 \text{ and } w > 0.$$

The gross rate of return to capital, R, is constant, as is the average and marginal product of labor, w. Capital depreciates completely in one period, so the net rate of return to capital, the real interest rate, is constant, equal to $R-1$ and positive.[3] The competitive (before-tax) wage rate equals w.

A new generation of identical consumers is born in each period. Each consumer lives for two periods, labeled 1 and 2. A young consumer has preferences described by the linear utility function

$$(2)\quad u(c,l,d,m) = D(c-\lambda l) + d - \mu m,$$
$$\lambda > 0, \quad \mu > 0 \quad \text{and} \quad D > 0,$$

where c and d are consumption when young and old, l and m are labor supply when young and old, λ and μ are the constant marginal rates of substitution between leisure and consumption when young and old, and D is the gross rate of time preference (equal to one plus the rate of time preference). Utility is normalized so that the marginal utility of consumption in period 2 is unity.

---

[2] Since writing this paper, we have become aware of ideas that resemble our own in the work by David Kreps (1984) on "corporate cultures," and by Jacques Crémer (1986) on cooperation in ongoing organizations. There is also a certain parallel between our suggested mechanism and the mechanism known as "posting a bond" in the labor economies literature. In both cases a cheater forfeits some resources by cheating. In our case the mechanism does not need any outside enforcement (by the courts) but is instead self-enforcing.

[3] An alternative interpretation of the model, which leads to equivalent results (since we disregard any nonnegativity conditions on the capital stock), is that the country is a small open economy facing a perfect world capital market with a constant (real) interest rate equal to $R-1$.

The feasible set of consumption and labor supply is given by

$$(3) \qquad 0 \le \underline{c} \le c, \quad 0 < \underline{l} \le l \le \bar{l},$$
$$0 \le \underline{d} \le d, \quad \text{and} \quad 0 < \underline{m} \le m \le \bar{m}.$$

That is, there are nonnegative lower limits $\underline{c}$ and $\underline{d}$ on consumption when young and old, and positive upper and lower limits on labor supply when young and old.

There are taxes on labor in periods 1 and 2 and on savings. A young consumer faces budget constraints

$$(4) \qquad c + s = (w - \tau)l, \; s \ge 0, \quad \text{and}$$

$$(5) \qquad d = (R - \theta)s + (w - \sigma)m,$$

where s is savings, and τ, σ, and θ are absolute taxes on labor when young, labor when old, and savings. By assumption borrowing is not allowed.[4] Therefore savings is restricted to be nonnegative.

Given the linear utility function, it is easy to derive saving and labor decisions. Maximizing the utility function subject to the feasibility set and the budget constraints gives

$$(6) \qquad s = \begin{cases} (w - \tau)l - \underline{c} & \text{for } R - \theta \ge D \text{ and} \\ 0 & \text{for } R - \theta < D, \end{cases}$$

$$(7) \qquad l = \begin{cases} \bar{l} & \text{for } (w-\tau)\max((R-\theta)/D, 1) \ge \lambda \text{ and} \\ \underline{l} & \text{for } (w-\tau)\max((R-\theta)/D, 1) < \lambda, \end{cases}$$

$$(8) \qquad m = \begin{cases} \bar{m} & \text{for } w - \sigma \ge \mu \text{ and} \\ \underline{m} & \text{for } w - \sigma < \mu. \end{cases}$$

What matters for the saving decision is only the relation between the after-tax rate of return, $R - \theta - 1$, and the rate of time preference, $D - 1$. When the after-tax rate of return is above the rate of time preference, the consumer consumes the minimum amount $\underline{c}$ when young and saves the maximum amount. The maximum saving is the difference between his after-tax wage $(w - \tau)l$, which depends on the labor supply when young according to (7), and the minimum consumption. When the after-tax rate of return is below the rate of time preference, saving is equal to its minimum, which is zero since we assume no borrowing. When the after-tax return equals the rate of time preference, saving is indeterminate. In this case we assume that the consumer saves the maximum amount.

Similarly, what matters for the labor supplies is only the relation between the after-tax wage rate and the rate of substitution between consumption and leisure. With regard to labor supply in the first period, (7), one has to take into account that there are two potentially relevant marginal rates of substitution. When $R - \theta > D$, consumption in the first period is constant and at its minimum. Then the relevant marginal rate of substitution is between leisure in the first period and consumption in the second period, $D\lambda$, and the relevant after-tax wage rate is $(R - \theta)(w - \tau)$, measured in second-period goods. If instead $R - \theta < D$, saving is zero and consumption in the first period equals after-tax income in the first period. Then the relevant marginal rate of substitution is λ, and the relevant after-tax wage rate is $w - \tau$. This combined gives (7). If the after-tax wage rate is above the marginal rate of substitution, the best thing is to work the maximum hours. If the after-tax wage rate is below the rate of substitution, the best thing is to work the minimum hours. If the after-tax wage rate is equal to the rate of substitution, the consumer is indifferent between working as much as possible, and as little as possible, or anything in between. We assume in this case that he works the maximum hours. With regard to second-period labor supply in (8), $w - \sigma$ and μ is the relevant after-tax wage rate and marginal rate of substitution, both when $R - \theta > D$ and when $R - \theta < D$, since in the latter case saving is zero and second-period consumption is not necessarily at its minimum.

---

[4] Borrowing could be allowed, without affecting the qualitative results.

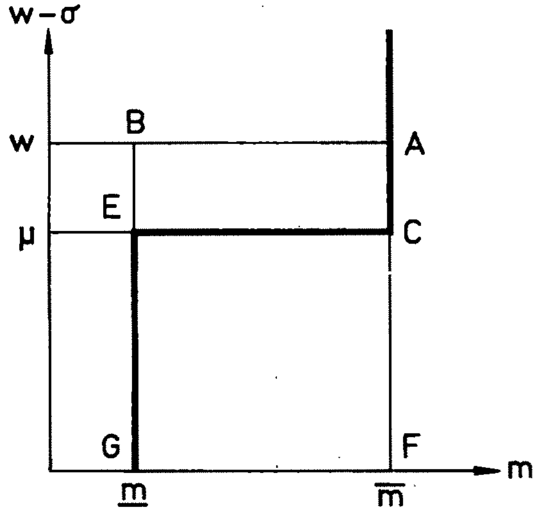


FIGURE 1

We also assume that the below-tax wage rate is above the rates of substitution between leisure and consumption and that the before-tax return to capital is above the rate of time preference

$$w > \lambda, \quad w > \mu, \quad \text{and} \quad R > D. \tag{9}$$

In the absence of taxes the consumer would work the maximum hours when young and old, and he would save the maximum amount when young.

Figure 1 further illustrates consumer behavior. The bold line shows the compensated (and uncompensated, since they are identical here) supply of labor when old, m, as a function of the net wage rate, $w-\sigma$. The before-tax wage rate w is above μ by assumption, and, in the absence of a tax on labor, the consumer when old would be in equilibrium at point A in the diagram. For sufficiently high taxes on labor, the net wage would fall below μ, and the consumer would work the minimum hours and be in equilibrium at some point on the vertical line EG. An equilibrium at the minimum hours is inefficient. The cost to the consumer of working $\overline{m}$ rather than $\underline{m}$ is the area $CEGF$, equal to $\mu(\overline{m}-\underline{m})$. But the value to society is the area $ABGF$, equal to $w(\overline{m}-\underline{m})$. The deadweight loss is the area $ABEC$, equal to $(w-\mu)(\overline{m}-\underline{m})$.

II. The First Best, the Second Best, and the Third Best

Below we consider equilibria when generations interact. In this section, however, we examine equilibria when each generation is economically independent. As mentioned, the representative body of a generation is called the council. In each period, there are two councils, one representing the young and one representing the old. The council of a given generation supplies a fixed amount of public goods, g, to its generation when the generation is old and can finance this second-period expenditure by taxing its generation in both periods. Before discussing equilibria with distortionary taxation, we describe the first-best optimum that would result if the council had access to lump-sum taxes. Let the council impose a second-period lump-sum tax equal to g on the old and set all other taxes equal to zero. Then in the first-best optimum with lump-sum taxation (denoted with an asterisk (*) superscript) labor supply when young, labor supply when old, and saving are at their maxima. Thus

$$l^* = \bar{l}, \quad m^* = \overline{m} \quad \text{and} \quad s^* = w\bar{l} - \underline{c}. \tag{10}$$

Consumption when young is at the minimum. Consumption when old is equal to savings plus wages when old minus the lump-sum tax. Hence, the utility level for the generation is

$$u^* = D(\underline{c} - \lambda\bar{l}) + Rs^* + w\overline{m} - g - \mu\overline{m}. \tag{11}$$

We use this equilibrium as a reference case in discussing equilibria with distortionary taxes.

Let us now assume that the council does not have access to lump-sum taxes. Instead it has the right to tax labor and savings of its generation as well as to borrow or lend. The council faces budget constraints

$$b = \tau l \quad \text{and} \tag{12}$$

$$g = \theta s + \sigma m + Rb, \tag{13}$$

where b is the council's saving in period 1. The two constraints can be collapsed into one intertemporal budget constraint

$$g = R\tau l + \theta s + \sigma m. \tag{14}$$

Consider first the case in which the council can make a binding commitment in period 1 to given taxes in period 2. As suggested by the discussion of Figure 1, for sufficiently low taxes there is no distortion of the labor supply and saving decisions. We assume that the maximum nondistortionary taxes on labor and savings fall short of the required public expenditure

$$g > R(w-\lambda)\bar{l} + (w-\mu)\bar{m} + (R-D)\bar{s}. \tag{15}$$

The maximum nondistortionary taxes are $\tau = w-\lambda$ and $\sigma = w-\mu$ on labor in period 1 and 2, and $\theta = R-D$ on savings. The maximum amount of savings, s, when the consumer faces the maximum nondistortionary tax on labor in period 1, $\bar{s}$, is $\bar{s} \equiv \lambda\bar{l} - \underline{c}$.

Without (15) the optimal taxation problem would be trivial. With (15), the council is forced to increase one or several taxes above their maximum nondistortionary levels. If the tax on labor supply in period 2 is increased, labor supply drops to its minimum, causing a deadweight loss equal to $(w-\mu)(\bar{m}-\underline{m})$. If the tax on savings is increased, saving drops to zero, causing a deadweight loss equal to $(R-D)\bar{s}$. If the tax on labor in period 1 is increased, first-period labor supply drops to its minimum, causing a direct deadweight loss of $D(w-\lambda)(\bar{l}-\underline{l})$ plus an indirect deadweight loss—due to lower savings—of $(R-D)(\bar{s}-\underline{s})$, where $\underline{s}$ is given by $\underline{s} \equiv (w-\tau)\underline{l} - \underline{c} \leq \lambda\underline{l} - \underline{c}$.[5] We now assume that the deadweight loss from distorting period-2 labor is smaller than the other distortions:

$$(w-\mu)(\bar{m}-\underline{m}) < \min\left[D(w-\lambda)(\bar{l}-\underline{l}) + (R-D)(\bar{s}-\underline{s}), (R-D)\bar{s}\right]. \tag{16}$$

This means that the council prefers to increase taxes on labor in period 2, rather than to increase taxes on savings or period-1 labor. We next assume that increasing taxes on period-2 labor above the distortionary level provides enough revenue to finance public expenditure. We later make the stronger assumption that distortionary taxation of labor in period 2 raises enough revenue to pay for g in the case when no saving is forthcoming; hence

$$g < R(w-\lambda)\bar{l} + w\underline{m}. \tag{17}$$

Given the assumptions (15)–(17), labor supplies and saving in the second-best optimum under commitment are

$$l = \bar{l}, \quad m = \underline{m}, \quad \text{and} \quad s = \bar{s}. \tag{18}$$

The tax rates (denoted by an overbar) are

$$\bar{\tau} = w-\lambda, \quad \bar{\theta} = R-D, \quad \text{and} \quad \bar{\sigma} = (g - R\bar{\tau}\bar{l} - \bar{\theta}\bar{s})/\underline{m}. \tag{19}$$

The only distortion is that labor in period 2 has dropped to its minimum. Consequently, the utility level for the generation is

$$\bar{u} = u^* - (w-\mu)(\bar{m}-\underline{m}). \tag{20}$$

It falls short of the first-best utility level by the deadweight loss from distorting labor supply in period 2.

---

[5] Notice that D multiplies $(w-\lambda)(\bar{l}-\underline{l})$. This is because utility is normalized to that the marginal utility of second-period consumption is unity.

The second-best optimum under commitment is unenforceable. There is nothing in the model that enforces the commitments to particular taxes in period 2. Indeed, since capital is predetermined in period 2, the council in period 2 has an incentive to deviate from the preannounced policy and levy a nondistortionary tax on capital if it can thereby avoid a distortionary tax on second-period labor. Such will be the case if, as we assume, the savings tax base, $R\bar{s}$, is sufficiently large to make it possible to lower labor taxes in period 2 to the nondistortionary level; that is,

$$(21)\quad g < R(w-\lambda)\bar{l} + (w-\mu)\overline{m} + R\bar{s}.$$

If (21) holds and capital equals $\bar{s}$, the council in period 2 would optimally set the tax on savings equal to

$$(22)\quad \hat{\theta} = [g - R(w-\lambda)\bar{l} - (w-\mu)\overline{m}]/\bar{s}.$$

By (15) it follows that $\hat{\theta}$ is large enough to reduce the net rate of return below the rate of time preference

$$(23)\qquad R - \hat{\theta} < D.$$

Equation (23) implies that a young consumer anticipating the council's behavior would not save (see (6)). The reduction of saving from $\bar{s}$ to zero produces a deadweight loss equal to $(R-D)\bar{s}$, and erodes the tax base in period 2. The council is left with no choice but to levy a tax on labor in period 2 above the distortionary level.

It follows that the third-best optimum without precommitment will have the following taxes (denoted by a caret(^))

$$(24)\qquad \hat{\tau} = w - \lambda = \bar{\tau} \quad \text{and}$$
$$\hat{\sigma} = (g - R\bar{\tau}\bar{l})/\underline{m} > \bar{\sigma},$$

in addition to the tax $\hat{\theta}$ on savings.[6] Note that our assumption (17) ensures that there exists a tax on labor in period 2 which fulfills (24). The labor and saving decisions yield

$$(25)\qquad l = \bar{l},\quad m = \underline{m} \quad \text{and} \quad s = 0.$$

The utility level is

$$(26)\qquad \hat{u} = \bar{u} - (R-D)\bar{s}$$
$$= u^* - (w-\mu)(\overline{m} - \underline{m})$$
$$- (R-D)\bar{s}.$$

Compared to the unenforceable second-best optimum, the third-best optimum has an additional deadweight loss, namely the saving distortion.

III. Enforcing the Second-Best Optimum Through a Social Contract

We now propose a mechanism, namely a social contract, that enforces an outcome very close to the second-best optimum. The social contract has the following two clauses:

(C1) Capital taxation (either explicit or implicit) above the prespecified level $\theta = R - D$ is prohibited.

(C2) The council of the young contribute a specified transfer Q to the council of the old.

We assume that each young generation can set up a social contract like (C1–2) at a transactions cost T. Notice that if the (councils of) succeeding generations stick to the social contract, each young generation effectively purchases the social contract from the coexisting old generation.

We now show that there exists an equilibrium with the properties that, starting from a society without a social contract, a social contract is set up by an initial generation and that, once set up, it is fulfilled by all subsequent generations. Furthermore, every council acts individually, rationally, and expectations are never falsified. In the language of game theory, the equilibrium is subgame perfect.

---

[6] This outcome would not result if levying distortionary taxes on first-period labor alone would yield revenue to finance g *and* less deadweight loss than taxing second-period labor and savings. To rule out this possibility, we need to assume either that $g > Rw\underline{l} + (w-\mu)(\overline{m} - \underline{m}) + (R-D)\underline{s}$ or that $D(w-\lambda)(\bar{l} - \underline{l}) + (R-D)(\bar{s} - \underline{s}) > (w-\mu)(\overline{m} - \underline{m}) + (R-D)\bar{s}$.

The intergenerational game has sequential moves in each period. The young decide whether or not to fulfill the existing social contract after the old have decided on the tax structure. We could think of the councils of the young and the old forming a joint government and bargaining about the price of the social contract, that is, about the transfer Q. We do not study the details of this bargaining process, but since we are dealing with a stationary environment it seems reasonable to treat Q as a constant. Our results below will give the possible range for the transfer, however.

To understand how the social contract comes into being, suppose the economy is initially contractless and consider the economic incentives to establish the social contract. It is useful to specify the extensive form of the game. Figure 2 shows part of the game tree. Node N_1 corresponds to an initial situation without a social contract in an arbitrary start-up period, called period 1. The council of the young in period 1, called the first generation, faces a choice between either setting up a social contract (move "S") or continuing without a social contract and remain in the *third*-best equilibrium (move "3"). Node N_2 corresponds to the beginning of the next period, called period 2, when the first generation in period 1 chose to set up a social contract and now has become old. At node N_2 the first generation faces a choice between either fulfilling clause (C1) of the social contract (move "F") or breaking the clause and overtaxing capital beyond the prespecified level (move "O"). Depending upon the choice of the first generation, node N_4 or N_5 is reached. If the first generation in period 1 chose to continue in the third-best equilibrium, node N_3 is reached in the beginning of period 2, and there is no choice for the first generation but to continue in the third-best equilibrium (3), and the society remains without a social contract.

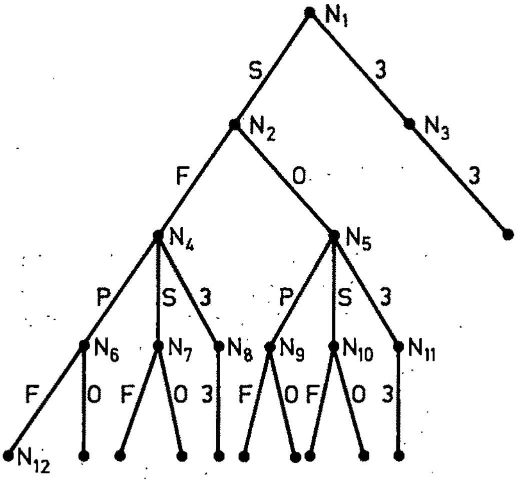


FIGURE 2

At nodes N_4 and N_5 the young in period 2, called the second generation, face three alternatives. They can choose to obey clause (C2) of the social contract and pay the transfer Q to the first-generation old (move "P"), they can set up a new social contract (S), or they can choose the third-best equilibrium (3). These choices lead to one of the nodes N_6–N_{11} in the beginning of period 3 when the second generation has become old. Now they face the choice between fulfilling (F) or breaking (O) clause (C1) of the social contract in nodes N_6, N_7, N_9, and N_{10}, whereas there is no choice but to continue in the third-best equilibrium in nodes N_8 and N_{11} (3).

Assume now that all generations (councils) follow an identical strategy, namely,

(S1) When young: (a) if a social contract exists and the existing contract has never been broken, fulfill clause (C2), that is, pay the transfer Q (move P), (b) otherwise, do not pay the transfer but set up a new social contract of the form (C1–2) (move S).

(S2) When old: (a) if an unbroken social contract exists of the form (C1–2), fulfill clause (C1), that is, do not overtax capital (move F), (b) otherwise, if a social contract exists, overtax capital (move O), or if a social contract does not exist, continue in the third-best equilibrium (move 3).

In the subsequent discussion we derive conditions which guarantee that no generation has an incentive to deviate unilaterally from the strategy (S1–2). This is exactly what is required in a subgame-perfect equilibrium.

We hence want to show that the path $N_1N_2N_4N_6N_{12}$ is a subgame-perfect equilibrium. Let us start with node N_1. The first generation of young to devise the social contract (move S) has the special advantage of not having to purchase the social contract, that is, to pay the transfer Q. On the other hand, it has to bear the transactions cost of establishing the social contract. The budget constraints confronting the council of the first generation are

$$b = \tau l \quad \text{and} \tag{27a}$$

$$g - Q = \theta s + \sigma m + Rb, \tag{27b}$$

where including Q in (27b) assumes that the second generation follows (S2). The budget constraints of each member of the first generation are given in (4) and (5). The transactions cost is T utils, incurred when young, but measured in second-period utility. We assume that the transactions cost for establishing a new social contract is the same for all generations irrespective of whether a social contract has existed or not.[7]

The first generation levies nondistortionary taxes θ^1 on savings and τ^1 on first-period labor supply, where

$$\theta^1 = R - D = \bar{\theta} \quad \text{and} \tag{28}$$

$$\tau^1 = w - \lambda = \bar{\tau}.$$

The corresponding tax on the first-generation's labor in period 2, σ^1, is distortionary and is given by

$$\sigma^1 = (g - Q - R\bar{\tau}\bar{l} - \bar{\theta}\bar{s})/\underline{m} < \bar{\sigma}.^8 \tag{29}$$

The labor and saving decisions resulting from this choice of taxes are

$$l = \bar{l}; \quad m = \underline{m}, \quad \text{and} \quad s = \bar{s}. \tag{30}$$

and the utility level of the first generation is thus

$$u^1 = \bar{u} + Q - T. \tag{31}$$

Note that this utility level is just the second-best level of utility less the transactions cost plus the value Q obtained from selling the social contract to the second generation.

The first generation of young will institute the social contract—that is, follow (S1b) and choose S at node N_1—if instituting the social contract increases their utility above the third-best optimum, that is, if

$$u^1 \geq \hat{u}. \tag{32a}$$

According to (26), (32a) requires

$$Q + (R - D)\bar{s} \geq T. \tag{32b}$$

Condition (32b) states that the gain from the sale of the social contract plus the reduction in the saving distortion from achieving a second-best tax structure must exceed the transactions cost of establishing the social contract.

We must also show that the first generation has an incentive to fulfill the social contract when old—that is, to follow (S2) and choose F at node N_2. If this generation fulfills the social contract when it is old, its utility when old is

$$v^1 = \bar{v} + Q, \tag{33}$$

---

[7] We could also let the transactions' cost be lower if a social contract already exists, without changing the qualitative results.

[8] This applies if $\sigma^1 > w - \mu$, in which case the second-period labor distortion remains. If $\sigma^1 < w - \mu$, the distortion is eliminated and the first generation is even better off.

where $\bar{v}$ is the period-2 utility in the unenforceable second-best optimum. Thus the gain in old age utility from sticking to the social contract relative to the unenforceable second best is Q, the receipt from the sale of the social contract. If the first generation breaks the social contract in their second period (O) it will impose the following taxes

$$(34) \qquad \tilde{\sigma}^1 = w - \mu \quad \text{and}$$

$$\tilde{\theta}^1 = [g - R(w-\lambda)\bar{l} - (w-\mu)\overline{m}]/\bar{s} = \hat{\theta}.$$

Second-period utility in this case, $\tilde{v}^1$, is $\bar{v}$ plus the elimination of the second-period labor distortion, $(w-\mu)(\overline{m}-\underline{m})$,

$$(35) \qquad \tilde{v}^1 = \bar{v} + (w-\mu)(\overline{m}-\underline{m}).$$

The necessary condition for the first generation to fulfill the social contract in period 2 is

$$(36a) \qquad v^1 \geq \tilde{v}^1,$$

which implies

$$(36b) \qquad Q \geq (w-\mu)(\overline{m}-\underline{m}).$$

This yields a lower bound on the price of the social contract.

When faced at node N_4 with the social contract set up by the first generation, the second generation, according to (S1), buys the social contract from the first generation. For this to hold the second generation must prefer purchasing the existing social contract (move P) to (1) not purchasing the existing social contract and establishing its own social contract (move S), and to (2) pursuing the third-best optimum (move 3). If the second generation's council purchases the existing social contract and is able to resell the social contract, its council faces the following budget constraints

$$(37a) \qquad Q + b = \tau l \quad \text{and}$$

$$(37b) \qquad g - Q = \theta s + \sigma m + Rb,$$

which together imply

$$(38) \quad g + (R-1)Q = R\tau l + \theta s + \sigma m.$$

Equation (38) points out that the second generation's council effectively rents the social contract for the amount $(R-1)Q$ since it purchases the social contract when young and resells it when old. Given (38) the optimal tax rates are the nondistortionary taxes on first-period labor supply and saving, $\bar{\tau}$ and $\bar{\theta}$, and the distortionary tax σ^2 on second-period labor supply.[9] The latter is determined by

$$(39) \quad \sigma^2 = [g + (R-1)Q - R\bar{\tau}\bar{l} - \bar{\theta}\bar{s}]/\overline{m}$$

$$> \bar{\sigma}.$$

Since the second generation achieves the unenforceable second-best optimum except that it has to rent the social contract, its utility u^2 is

$$(40) \qquad u^2 = \bar{u} - (R-1)Q.$$

For the second generation to be willing to purchase the social contract rather than create its own, u^2 must exceed u^1; hence,

$$(41a) \qquad u^2 \geq u^1.$$

If (41a) holds (32) ensures that u^2 exceeds the third-best optimum $\hat{u}$. Conditions (31), (40), and (41a) imply

$$(41b) \qquad T \geq RQ.$$

In words, (41b) requires that the cost of creating a new social contract exceed the gain from not having to purchase the old social contract. This condition yields an upper bound on the admissible price of the social contract.

The second generation must also prefer to fulfill the social contract when they are old; that is, prefer to follow (S2) and choose F at node N_6. Consequently, v^2, the second generation's utility when old if it purchases the social contract must exceed $\tilde{v}^2$, the utility when old if it reneges and overtaxes capital (move O). Compared to the unenforceable second best, the consumption of the second generation in period 2 if it obeys the social contract is smaller by the amount of rent on the social contract. Hence,

---

[9] We assume that taxing second-period labor is enough to finance *total* spending of the council: $g + (R-1)Q$. Thus there is no additional distortion caused by raising the revenue for the rent of the law $(R-1)Q$. In a nonlinear model, we would have to add an additional welfare triangle to the cost of acquiring the social contract, however.

$$v^2 = \bar{v} - (R-1)Q. \tag{42}$$

If the second generation breaks the social contract it foregoes the contribution Q from the third generation and imposes taxes according to

$$\tilde{\sigma}^2 = w - \mu \quad \text{and} \tag{43}$$

$$\tilde{\theta}^2 = \left[g + RQ - R(w-\lambda)\bar{l} + (w-\mu)\bar{m}\right]/\bar{s} > \hat{\theta}.$$

Each member of the second generation will respond to these second-period taxes by working full time, and their second-period utility will be

$$\tilde{v}^2 = \bar{v} - RQ + (w-\mu)(\bar{m} - \underline{m}). \tag{44}$$

The condition for preferring F to O is

$$v^2 \geq \tilde{v}^2, \quad \text{or} \tag{45a}$$

$$Q \geq (w-\mu)(\bar{m} - \underline{m}), \tag{45b}$$

which is identical to (36b).

Finally, we have to verify that the second generation would not deviate from (S1) if the first generation in fact did not set up a social contract when young or broke the social contract when old. Faced with no existing social contract, the second generation would have the same problem as the first. Consequently, if condition (32) holds, (S1b) is optimal. Faced with a broken social contract at node N_5, the second generation could contemplate deviating from (S1b) and (S) either by pursuing the third-best optimum (3), or by purchasing the social contract even though it had been broken (P). If (32) holds, setting up a new social contract is better than the third best. In contrast, purchasing the broken social contract would be worse than the third best because the social contract could not be resold to the third generation, given that the third generation follows (S1b), leaving the second generation with third-best utility $\hat{u}$ less RQ. The above argument shows that the threat of punishment facing the first generation is not empty. If the first generation breaks the social contract when old, it is indeed optimal for the second generation to carry out the punishment inherent in (S1b).

An alternative justification for (S1b) that does not require invoking punishment behavior, albeit selfish punishment behavior, by future generations is simply to view transactions costs in the following way. Since the old and young generations have formed a government, a decision of the old to break the social contract would impose transactions costs on the young who will have to participate in negotiations about the capital levy. Once the young see that they will have to bear the transactions cost of redrafting the existing social contract, they will decide they are better off setting up their own social contract. Thus transactions costs may be such that once a social contract/institution is broken, it is in effect destroyed and cannot be subsequently sold. In this case strategy (S1b) follows automatically since it is impossible to purchase a social contract/institution that no longer exists.

In summary, we have found that deviations from (S1–2) by the first and second generations are indeed ruled out by conditions (32), (36), and (41). Because the third and subsequent generations are in the same situation as the second generation, the conditions (32), (36), and (41) apply to them as well. We have therefore shown that, provided these conditions hold, there is a perfect equilibrium in which the social contract (C1–2) is instituted once and then sold from generation to generation. The conditions to be satisfied in such an equilibrium can be reexpressed as the inequalities

$$(46)\quad (w-\mu)(\overline{m}-\underline{m}) < T/R < (R-D)\bar{s}/(R-1) \quad \text{and}$$

$$(47)\quad \max[T-(R-D)\bar{s},(w-\mu)(\overline{m}-\underline{m})] < Q < T/R.$$

The transactions cost T must satisfy inequality (46). For given T, any Q in the interval (47) will do.

To obtain a sense of the potential range of Q satisfying these constraints consider the following parameter values

$$R=2,\quad D=1.8,\quad \bar{s}=1,\quad w=1,\quad \mu=.85,$$
$$\bar{l}=\overline{m}=1,\quad \underline{m}=.9,\quad \text{and}\quad T=.04.$$

Since one period is roughly 30 years of real time, setting R equal to 2 is equivalent to assuming an annual real interest rate slightly greater than 2 percent. According to these parameters pretax national income in both the unenforceable and the enforceable second-best steady states is 2.9, since the young earn 1, the old earn .9, and capital income is 1. Income in the third-best steady state is only 1.9 reflecting the loss of capital income. The labor supply distortion in the second period is .015, or slightly more than one-half percent of total second-best income. According to (46) T must exceed .03, which is slightly more than 1 percent of second-best income. This value of T may seem large, especially if one considers that income in this model corresponds roughly to income over a 30-year period. However, the costs of setting up complex institutions have, at times, been very substantial, including, in the extreme, costly wars to settle disputes. If T is assumed to equal .04, then Q must lie between .02 and .015.

IV. Enforcement Without Transactions Costs

A second-best equilibrium may be supported by social contracts also if there are no transactions costs. Suppose that there are no transactions cost and that each generation adopts the strategy consisting of (S2) and

(S1′) When young: (a) if a social contract exists and the existing contract has never been broken, obey clause (C2), that is, pay the transfer Q (move P), (b) if a social contract has ever been broken, do not pay the transfer and do not set up a new social contract (move 3), (c) otherwise, do not pay the transfer but set up a new social contract of the form (C1–2) (move S).

The new element is (b), which results in the third-best equilibrium forever if a social contract is ever broken.

It is straightforward to verify that this strategy supports a subgame-perfect equilibrium, provided that Q lies in the interval given by

$$(48)\quad (w-\mu)(\overline{m}-\underline{m}) < Q < (R-D)\bar{s}/(R-1).$$

The strategy (S1′–2) is in the nature of a James Friedman (1971) "trigger strategy," in that a deviating generation is punished by the succeeding generations reverting to the "stage-game equilibrium," which here is the third best.[10] By (S1b) there is something of a trigger strategy element also in our equilibrium with transaction costs. In the equilibrium with transactions costs, however, a deviating generation is not punished by the succeeding generation reverting to the "stage-game equilibrium." Instead it is punished by the succeeding generation's setting up a social contract of its own, which is a second-best and Pareto-optimal equilibrium given that lump-sum taxes are infeasible.[11] We believe that this feature makes our transactions-cost equilibrium much more plausible than the standard trigger strategy equilibrium.

---

[10] Trigger strategies abound in the macro literature on how reputation can enforce *ex ante* optimal policies. Despite the formal similarities, there are conceptual differences between the arguments. In the usual reputation argument the disincentive to deviate for an *infinite*-horizon policymaker is the threat of future policy failures. Here, the disincentive to deviate for a *finite*-horizon policymaker is instead the threat of an immediate capital loss on the policymaker's asset.

It is interesting that also in situations where there are no transactions costs there exist strategies with equilibria where the economy after a deviation quickly returns to the second-best situation. Consider the strategies consisting of (S2) and

(S1″) When young: (a) if a social contract exists, the existing contract has never been broken, and the existing contract was not set up by a generation who broke the previous contract, obey clause (C2), that is, pay the transfer Q (move P), (b) otherwise, do not pay the transfer but set up a new social contract of the form (C1–2) (move S).

The new element relative to (S1) is that the young generation sets up a new contract and does not pay to the old generation, if the old generation broke a previous social contract and set up their own social contract in order to avoid paying the transfer.[12] This provides punishment of young who choose S rather than P in node N_4, and yet restores the second-best equilibrium. It is straightforward to show that there exists a subgame-perfect equilibrium with strategy (S1″–2) for Q obeying the condition (48).

V. Extensions, Broader Implications, and Conclusion

One obvious question to raise at this stage is whether the sale of social contracts can enforce seemingly time-inconsistent taxation in other optimizing models. Among neoclassical intertemporal models Barro's (1974) altruistic, infinite horizon model is the principal alternative to the life-cycle model, so it may be useful to consider the sale of social contracts in that model. As is well known, the main distinction between the two models is that in contrast to the life-cycle model, intergenerational redistribution per se in Barro's model is of no economic consequence. Hence, current old generations will not care about the price at which they sell social contracts to the next generation since they know that changes in that price simply redistribute between themselves and their children. It appears that the sale of social contracts would be useless in the Barro model.

A second question about extensions of our model is whether it covers other types of taxes and other types of fiscal policy. The answer appears to be yes. There is no reason to preclude the writing of social contracts or the making of agreements that cover the entire array of feasible direct and indirect taxes on goods and assets. Social contracts or agreements could also extend to investment incentives, changes in which can lead to capital losses and thus constitute covert government asset taxation (see, for example, Alan Auerbach and Laurence Kotlikoff, 1987, ch. 9).

The sale of social contracts could also enforce Pareto-improving intergenerational transfers, such as unfunded Social Security, when the economy is beyond the Golden Rule level of capital accumulation. The enforcement problem here, of course, is that each young generation might choose to renege on the existing system and instead set up its own intergenerational transfer program. The same type of enforcement problem arises in Paul Samuelson's (1958) consumption loan model; in that model each young generation, in the absence of an enforcement mechanism, would choose to issue its own money rather than transfer to the current old by accepting their money.[13]

---

[11] Like most subgame-perfect equilibria, our transactions-cost equilibrium is not renegotiation proof, however. (See, for example, Joe Farell and Erik Maskin (1987) for a discussion of this equilibrium concept, which is a recent refinement of subgame perfection.) It fails on one point: When punishing a preceding deviating generation by setting up a new social contract in node N_5, the subsequent young generation incurs the transaction cost T, which is inefficient.

[12] This strategy is an example of Peter Hammond's (1975) trigger strategy: If the old were "noncooperative," the young should not transfer anything to the old and be labeled "cooperative," or be labeled noncooperative. If the old were cooperative, the young should give the transfer to the old and be labeled cooperative, or be labeled noncooperative. (We were alerted to the relation to Hammond's work by Merwan Engineer, 1986.)

While our model assumes that only the old consume the public good, it appears to generalize to include joint consumption of the public good by young and old. Payment from the young to the old may also occur under the guise of contributions to finance unfunded Social Security benefit payments or, more generally, by the young adopting debts of the old.[14]

A third question is whether the sale of social contracts can enforce time consistency of policies other than fiscal policies in life-cycle type models. In particular, what about monetary policy? Again the answer appears to be yes. In principle, it should be possible to solve the surprise-inflation problem studied by Leonardo Auernheimer (1974) and Guillermo Calvo (1978) with a social contract or agreement or institution determining the rate of money creation. If such a social contract is sold to the next generation, the temptation to run a surprise inflation may be outweighed by the potential loss of revenues from the sale of the anti-inflation social contract. Taking the model in this paper literally, one may of course argue that intergenerational aspects are of minor importance when it comes to the short-run conduct of monetary policy. However, it may be that similar mechanisms can work when individual policymakers with finite horizons are appointed to overlapping terms of office in a long-lived committee, such as in the Board of Governors of the Federal Reserve.[15]

Finally, the notion of selling agreements to enforce time consistency may have wider applications. Take as an example seniority-based wage scales in unionized firms. In establishing unions initial older union workers may institute a steep wage-tenure profile to maximize their own lifetime earnings. Steep wage-tenure profiles can reward those with substantial tenure and avoid sharing sizable amounts of rents with new union hires who are willing to take less than they can earn elsewhere when young in exchange for higher wages in middle and old age. Time-consistency problems may arise in such union contracts. Suppose first, there are young, middle age, and older union members; second, that the union is controlled by the oldest union members; and third, that the only other group capable of reestablishing a union are the middle-age workers. If a steep wage tenure profile is established, the oldest union members will, at any point in time, have an incentive to renege on their promised pay increases to middle-age members and simply hire (or have the firm hire) more lower-paid young workers. Breaking their social contract in this manner may, however, reduce or eliminate their ability to attract new workers. Since the older workers' payments in excess of their marginal product may be financed, in large part, by paying younger workers less than their marginal product, tearing up the union agreement by reneging on pay increases to middle-age workers may be too costly. The other group that could break the union agreement is current middle-age workers. If they could costlessly toss out current old members and garner the excess wage payments to the oldest members for themselves by setting up a new union they would. However, if the transactions cost of such action exceeds the gain a new union will not occur.[16]

---

[13] Engineer (1986) adopts our framework to show how transfer institutions such as money may be enforced in Samuelson's consumption loan model.

[14] The intergenerational transfer associated with the sale of laws will also have an independent effect on savings, reducing savings in the sale of laws equilibrium below that in the unenforceable second best. While third-best savings in our model is zero, it seems possible that, in a less stylized model, savings in the third best could exceed savings in the sale of laws equilibrium.

[15] Guido Tabellini (1987) and Douglas Waldo (1987) show that a low inflation rule may be time consistent when policymakers with finite horizons overlap in a committee like the Board of Governors, which decides on monetary policy by majority vote. What prevents deviation from the low inflation rule in these papers is not an "intergenerational" transfer scheme between the committee members, however, but rather that the "median voter" in the committee does not have a last period of office, unlike an individual finite-horizon policy-maker.

To conclude, the sale of social contracts, agreements, and, more broadly, institutions appears to provide *ex post* enforcement of behavior preferred *ex ante* in a range of public and private contexts. These arrangements need not be explicit, and the payment for their purchase may be made in a variety of forms. As a consequence it may prove difficult to refute empirically this subtle resolution of the time-consistency problem.

---

[16] Alberto Alesina and Stephen Spear (1987) exploit the ideas in our paper and in Crémer (1986) in yet another context. They show how conflicts of interest between senior and junior members of a political party may be resolved by help of a transfer scheme within the party.

REFERENCES

Alesina, Alberto and Spear, Stephen E., "An Overlapping Generations Model of Electoral Competition," mimeo., 1987.

Auerbach, Alan and Kotlikoff, Laurence J., *Dynamic Fiscal Policy*, New York: Cambridge University Press, 1987.

Auernheimer, Leonardo, "The Honest Government Guide to the Revenue from the Creation of Money," *Journal of Political Economy*, May/June 1974, *82*, 598–606.

Backus, David and Driffill, John, "Inflation and Reputation," *American Economic Review*, June 1985, *75*, 530–538.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, May/June 1974, *82*, 1095–117.

______ **and Gordon, David B.,** "Rules, Discretion and Reputation in a Model of Monetary Policy," *Journal of Monetary Economics*, April 1983, *12*, 101–21.

Calvo, Guillermo A., "On the Time Consistency of Optimal Policy in A Monetary Economy," *Econometrica*, November 1978, *46*, 1411–28.

Crémer, Jacques, "Cooperation in Ongoing Organizations," *Quarterly Journal of Economics*, February 1986, *101*, 33–50.

Eaton, Jonathan, "Lending with Costly Enforcement of Repayment and Potential Default," mimeo., 1985.

Engineer, Merwan, "Costly Transfer Institutions and Cooperation Between the Generations," mimeo., Queens University, 1986.

Farell, Joe and Maskin, Eric, "Notes on Renegotiation in Repeated Games," mimeo., 1987.

Fischer, Stanley, "Dynamic Inconsistency, Cooperation and the Benevolent Dissembling Government," *Journal of Economic Dynamics and Control*, February 1980, *2*, 93–107.

Friedman, James W., "A Non-cooperative Equilibrium for Supergames," *Review of Economic Studies*, January 1971, *38*, 1–12.

Hammond, Peter, "Pure Exchange Equilibrium of Dynamic Economic Models," in E. S. Phelps, ed., *Charity: Altruism or Cooperative Egoism*, Russell Sage Foundation, 1975.

Kreps, David M., "Corporate Cultures and Economic Theory," mimeo., 1984, Stanford University.

Kydland, Finn E. and Prescott, Edward C., "Rules Rather than Discretion: The Inconsistency of Optimal Plans," *Journal of Political Economy*, June 1977, *85*, 473–92.

Lucas, Robert E., Jr. and Stokey, Nancy L., "Optimal Fiscal and Monetary Policy in an Economy Without Capital," *Journal of Monetary Economics*, April 1983, *12*, 55–93.

Persson, Torsten, "Credibility of Macroeconomic Policy: An Introduction and a Broad Survey," *European Economic Review*, March 1988, *32*, 519–32.

Persson, Mats, Persson, Torsten and Svensson, Lars E. O., "Time-Consistency of Fiscal and Monetary Policy," *Econometrica*, November 1987, *55*, 1419–31.

Rogoff, Kenneth, "Reputational Constraints on Monetary Policy," *Carnegie-Rochester Conference Series on Public Policy*, Spring 1987, *26*, 141–81.

Samuelson, Paul R., "An Exact Consumption-Loan Model of Interest with or without the Social Contrivance of Money," *Journal of Political Economy*, December 1958, *66*, 467–82.

Tabellini, Guido, "Reputational Constraints on Monetary Policy: A Comment," *Carnegie-Rochester Conference Series on Public Policy*, Spring 1987, *26*, 183–90.

Waldo, Douglas G., "Committees, Staggered Terms and Unraveling," mimeo., 1987.

Innovation in Large and Small Firms: An Empirical Analysis

By ZOLTAN J. ACS AND DAVID B. AUDRETSCH*

We present a model suggesting that innovative output is influenced by R&D and market structure characteristics. Based on a new and direct measure of innovation, we find that (1) the total number of innovations is negatively related to concentration and unionization, and positively related to R&D, skilled labor, and the degree to which large firms comprise the industry; and (2) these determinants have disparate effects on large and small firms.

As Simon Kuznets (1962) observed, perhaps the greatest obstacle to understanding the role of innovation in economic processes has been the lack of meaningful measures of innovative inputs and outputs. More recently, there has been the development of new data sources measuring different aspects of technical change. These new sources of data have included measures of patented inventions from the computerization by the U.S. Patent Office (Bronwyn Hall et al., 1986; Adam B. Jaffe, 1986; Ariel Pakes and Zvi Griliches, 1980), better measures of research and development (John Bound et al., 1984, and F. M. Scherer, 1982), and stock market values of inventive output (Pakes, 1985). While several of these new and improved data sources have been used to examine the relationship between innovative activity and firm size, there have been virtually no studies able to apply a more direct measure of the innovative output. For example, the limitations of using patent data were significant enough to supplement them with renewal data (Pakes and Mark Schankerman, 1984). Further, while most of the empirical research has examined only the innovative activity contributed by relatively large firms, the innovative output of the smallest firms has received only scant attention and quantification.[1] Thus, most of the inferences which have been made about the causes of innovative activity have been based on observing only the behavior of larger firms.[2] Such inferences may be misleading since, as we show, almost half of the number of innovations are contributed by firms which employ fewer than 500 workers.

The purpose of this paper is to add to the literature on new measures examining technical change by introducing a more direct measure of innovative activity, to determine some of its basic properties, and to illustrate its use with a reduced form empirical model. We present a model which investigates the degree to which innovative output is affected by different industry characteristics, and the extent to which small and large firms respond differently to various stimuli. The econometric analysis enables the testing of

---

*Research Fellows, Wissenschaftszentrum Berlin, Reichpietschufer 50, D-1000 Berlin 30, Federal Republic of Germany. We wish to thank George Bittlingmayer, Bo Carlsson, Paul Geroski, Albert N. Link, Richard R. Nelson, William K. Scheirer, Joachim Schwalbach, J.-Matthias Graf von der Schulenburg, Hideki Yamawaki, Klaus Zimmermann, two anonymous referees, and seminar participants at the U.S. Small Business Administration, Case Western Reserve University, and the University of Bradford for helpful comments. We are especially grateful for the suggestions by F. M. Scherer and the computational assistance of Michael Karge and Jianping Yang. All errors and omissions remain our responsibility.

[1] For a thorough review of the literature relating technical change to innovation activity, see Morton I. Kamien and Nancy L. Schwartz (1975), F. M. Scherer (1980), and Richard C. Levin et al. (1985).

[2] For example, Scherer (1965) related market structure to the number of patents for fewer than 500 of the largest U.S. corporations.

two hypotheses: (1) the degree to which *R&D* expenditures produce innovative output is conditioned by the market structure characteristics; and (2) as Sidney G. Winter (1984) suggests, small- and large-firm innovative activity respond to distinct technological and economic regimes.

In Section I of this paper, we introduce the new innovation data and compare them with the more traditional measures of technical change, including patented inventions and *R&D* expenditures. After presenting the model in Section II, cross-section regressions estimating the logs of the 1982 total number of innovations, large-firm innovations, small-firm innovations, and the small-firm innovation share for 247 four-digit SIC (Standard Industrial Classification) industries are presented in Section III. Finally, in Section IV, a summary and conclusion of the model and empirical results are provided. We find that innovative output increases with industry *R&D* expenditures at a less than proportional rate. While some of the appropriability measures, such as market concentration and unionization, are negatively related to innovation activity, the extent to which an industry is comprised of large firms is positively related to the total number of innovations. Similarly, we find considerable evidence supporting the hypothesis by Winter (1984) that innovation activity for small firms responds to a different technological and economic environment than does innovation activity for large firms.

I. The Innovation Data

The measure of innovative activity used in this paper is the number of innovations in each four-digit SIC industry recorded in 1982. The data, which were only recently released by the U.S. Small Business Administration,[3] consist of 8,074 innovations introduced into the United States in 1982. Of these innovations, 4,476 were identified as occurring in manufacturing industries. The Small Business Administration constructed this data base by examining over 100 technology, engineering, and trade journals, covering each manufacturing industry. From the sections in each trade journal listing innovations and new products, a data base consisting of the innovations by four-digit SIC industry was formed. The entire list of trade journals used to compile these data is available from the authors. The Small Business Administration defines an innovation as "a process that begins with an invention, proceeds with the development of the invention, and results in the introduction of a new product, process or service to the marketplace." (Keith L. Edwards and Theodore J. Gordon, 1984, p. 1) Because the innovations recorded in 1982 were the result of inventions made, on average, 4.2 years earlier, in some sense the innovation data base represents the inventions made around 1978 that were subsequently introduced to the market in 1982. The data were also checked for duplication. In fact, 8,800 innovations were actually recorded, but it was subsequently found that 726 of them appeared either in separate issues of the same journal or else in different journals. Thus, double counting was avoided. While the Small Business Administration does not claim that every single innovative activity is included in the data base, it does consider the data to be "comprehensive and reliable."

The innovation data were classified according to the industry of origin based on the SIC code of the innovating enterprise. The data were then classified into innovations by large firms, defined as firms with at least 500 employees, and innovations by small firms, defined as firms with fewer than 500 employees. For example, an innovation made by a subsidiary of a diversified firm would be classified by industry according to the SIC industry of the innovating subsidiary (enterprise) and not by the SIC industry of the parent firm. However, the innovation would be classified by size according to the size of the entire firm and not just by the size of the subsidiary.[4]

---

[3] For a more complete explanation of the data, see the Appendix.

[4] Because 67 innovations could not be classified according to firm size, the number of total innovations does not always equal the sum of large- and small-firm innovations.

TABLE 1—NUMBER OF INNOVATIONS AND THE RELATIVE INNOVATIVE ACTIVITY OF LARGE AND SMALL FIRMS IN THE MOST INNOVATIVE INDUSTRIES, 1982

Industry	Total Innovations	Large-Firm Innovations/ Small-Firm Innovations	Large Firm Employment/ Small-Firm Employment
Electronic Computing Equipment	395	0.696	16.544
Process Control Instruments	165	0.731	1.632
Radio and TV Communication Equipment	157	1.153	4.814
Pharmaceutical Preparations	133	9.231	19.408
Electronic Components	128	0.740	0.894
Engineering and Scientific Instruments	126	0.518	1.096
Semiconductors	122	3.138	5.757
Plastics Products	107	0.268	0.332
Photographic Equipment	88	8.778	18.231
Office Machinery	77	6.700	11.658
Instruments to Measure Electricity	77	0.596	2.077
Surgical Appliances and Supplies	67	4.154	3.566
General Industrial Machinery	67	4.154	0.976
Surgical and Medical Instruments	66	0.833	3.566
Special Industry Machinery	64	2.048	2.717
Industrial Controls	61	0.326	1.732
Toilet Preparations	59	2.278	5.289
Valves and Pipe Fittings	54	0.606	2.436
Electric Housewares and Fans	53	7.833	17.182
Measuring and Controlling Devices	52	0.067	0.866
Food Products Machinery	50	3.083	1.976
Motors and Generators	49	3.900	14.625
Plastic Materials and Resins	45	2.000	6.937
Industrial Inorganic Chemicals	40	4.000	4.917
Radio and TV Receiving Sets	40	8.750	13.925
Hand and Edge Tools	39	2.455	2.125
Fabricated Platework	38	3.222	1.597
Fabricated Metal Products	35	0.706	0.495
Pumps and Pumping Equipment	34	1.125	3.484
Optical Instruments and Lenses	34	0.571	1.179
Polishes and Sanitation Goods	33	0.684	2.937
Industrial Trucks and Tractors	33	0.650	6.404
Medicinals and Botanicals	32	5.400	4.376
Aircraft	32	31.000	111.111
Environmental Controls	32	2.200	3.566
Total	2,617	1.272	8.440

Source: U.S. Small Business Administration.

There are several other important qualifications which should be made concerning the innovation data. The trade journals report relatively few process, service, and management innovations and tend to capture mainly product innovations. The most likely effect of this bias is to underestimate the number of innovations emanating from large firms, since larger enterprises tend to produce more process innovations than do their smaller counterparts. However, because it was found that the large-firm innovations are more likely to be reported in trade journals than are small-firm innovations, the Small Business Administration considers the biases to be at least somewhat offsetting.

Table 1 lists the total number of innovations in those industries which had the greatest number of innovations in 1982, along with the corresponding ratios of large-firm innovation to small-firm innovation and large-firm employment to small-firm employment (also for 1982). Large firms are defined as firms with at least 500 employees and small firms are defined as firms with fewer than 500. In some industries, the large firms exhibit considerably more innovative activity than do their smaller counterparts, while in other industries the small firms are apparently more innovative. One potential concern might be that the significance and "quality" of the innovations vary considerably between large and small firms. Based on 4,938 of the innovations, the Small Business Administration classified each innovation according to one of the following levels of significance: (1) the innovation established an entirely new category of product; (2) the innovation is the first of its type on the market in a product category already in existence; (3) the innovation represents a significant improvement in existing technology; and (4) the innovation is a modest improvement designed to update an existing product. While none of the innovations in the sample were in the highest level of significance, 80 were in the second level, 576 in the third level, and 4,282 were classified in the fourth level. Within each level of significance, the distribution between large- and small-firm innovations proved to be remarkably constant. In both the second and third significance categories, the large firms accounted for 62.5 percent of the innovations and the small firms for the remaining 37.5 percent. In the fourth significance category, the large firms accounted for just a slightly smaller share of the innovations, 56.6 percent, while the small firms contributed the remaining 43.4 percent. A chi-square test for the hypothesis that there is no difference in the frequency of innovation with respect to innovation significance and firm size cannot be rejected at the 99 percent level of confidence (Edwards and Gordon, 1984).[5] Thus, based on the U.S. Small Business Administration's classification of the significance level of innovations, there does not appear to be a great difference in the "quality" and significance of the innovations between large and small firms. However, the extent of innovative activity does not necessarily correspond to the market values of the innovations. It is conceivable that larger firms may tend to focus on innovations with a higher market value.

The second column of Table 1 shows that while the number of large-firm innovations exceeded that of small firms in twenty-one of the thirty-five most innovative industries, the small firms were more innovative in the remaining fourteen. By contrast, the third column indicates that large firms accounted for a greater share of 1982 industry employment in all but five of these highly innovative industries. For the entire sample of manufacturing industries the simple correlation between these two measures is only 0.316. While the ratio of the innovations-per-employee is 6.64 times greater for small firms than for large firms for the most innovative industries, small firms were only 43 percent more innovative than their larger counterparts for the entire sample of manufacturing industries.[6]

Table 2 shows the distribution of manufacturing industries according to innovation frequency. About one-quarter of the innovating industries contributed at least sixteen innovations, while slightly more than one-half of these industries had fewer than six innovations.

In order to compare the new innovation measures with the more traditional measures of technical change, Table 3 provides a correlation matrix of input and output measures of innovative activity. There is a striking difference in the simple correlation of 0.481

---

[5] Unfortunately, no further information regarding the distribution of innovations by significance category has been released by the U.S. Small Business Administration. Therefore, no comparisons across industries can be made.

[6] It should be noted that while there were a total of 8,074 innovations identified in the data base, some innovations occurred outside of manufacturing and others were contributed from firms which could not be found in publishing company directories. While some of these were subsidiaries of large firms, many more were small companies. A random sample of 600 firms (with 375 responses) was used to allocate the entire set of innovations into 55 percent from small firms and 45 percent from large firms, resulting in an innovation-per-employee ratio 2.38 times greater in small firms than in large firms.

TABLE 2—THE DISTRIBUTION OF MANUFACTURING INDUSTRIES ACCORDING TO INNOVATION FREQUENCY

	Number of Innovations[a]							
	1–5	6–10	11–15	16–20	21–25	26–30	31 >	Total
Total Innovations	149	36	21	13	15	6	36	276
(percent)	(53.99)	(13.04)	(7.61)	(4.71)	(5.43)	(2.17)	(13.04)	(99.99)
Large-Firm Innovations	127	28	24	6	7	8	18	218
(percent)	(58.26)	(12.84)	(11.01)	(2.75)	(3.21)	(3.67)	(8.26)	(100.00)
Small-Firm Innovations	129	33	16	9	4	1	13	205
(percent)	(62.93)	(16.10)	(7.80)	(4.39)	(1.95)	(0.49)	(6.34)	(100.00)

Source: U.S. Small Business Administration.

[a] Only industries in which there is at least some innovation activity are included in this table.

TABLE 3—CORRELATION MATRIX OF INPUT AND OUTPUT MEASURES OF INNOVATIVE ACTIVITY

	Total Innovations	Large-Firm Innovations	Small-Firm Innovations	Total *R&D* Expenditures	Company *R&D* Expenditures
Large-Firm Innovations	0.920	–			
Small-Firm Innovations	0.922	0.698	–		
Total *R&D* Expenditures	0.481	0.532	0.379	–	
Company *R&D* Expenditures	0.746	0.737	0.672	0.764	–
Patents	0.467	0.482	0.382	0.327	0.440

between total *R&D* expenditures (also from the 1977 FTC Line of Business) and total innovations, and of 0.746 between company *R&D* expenditures and total innovations. That the correlation between company *R&D* and innovative activity is stronger than the correlation between total *R&D* and innovative activity is consistent with the findings by Griliches (1986) that privately financed *R&D* has a larger effect on private productivity and profitability than does government-financed *R&D*. The somewhat lower correlation between the *R&D* measures and the total number of patented inventions[7] between 1970 and 1972 is not surprising, as Pakes and Griliches (1980, p. 378) observe that "patents are a flawed measure (of innovative output); particularly since not all new innovations are patented and since patents differ greatly in their economic impact." Still, the correlation between the output measures, total innovations, and patents is greater than that between the *R&D* measures and patents. Also of interest is that, just as the correlation between innovations and *R&D* is greater for large firms than for small ones, the correlation between patents and innovations is greater for large firms than for small ones.

The relationships between the input and output measures of innovative activity are somewhat sensitive to the total amount of innovative activity. That is, in industries in which there is little innovative activity, defined as those with fewer than seven innovations, the correlation between all of the measures of technical change become considerably weaker. In particular, the correla-

---

[7] The correlations involving the patents are at the three-digit SIC level. (The corresponding variables were also aggregated to the three-digit SIC level.) All other correlations are for the four-digit SIC level.

TABLE 4—COMPARISON OF INNOVATION DATA WITH *R&D* AND PATENT MEASURES[a]

Industry Group	Total Innovations	Patents	Patents/ Innovation	Company *R&D* (millions)	*R&D* (millions)/ Innovation
Food and Tobacco	206	311	1.51	272	1.32
Textiles and Apparel	29	147	5.07	65	2.24
Lumber and Furniture	83	50	0.60	37	0.45
Paper	61	292	4.79	150	2.46
Chemicals (excluding drugs)	332	3,492	10.52	1,260	3.80
Drugs	170	868	5.11	449	2.64
Petroleum	24	1,046	43.58	360	15.00
Rubber and Plastics	129	637	4.94	287	2.22
Stone, Clay, and Glass	59	477	8.09	149	2.53
Primary Metals	74	424	5.73	239	3.23
Fabricated Metal Products	340	450	1.32	246	0.72
Machinery (excluding office)	612	1,657	2.71	852	1.39
Computers and Office Equipment	566	1,045	1.85	1,054	1.86
Industrial Electrical Equipment[b]	444	836	1.88	210	0.47
Household Appliances	64	232	3.63	78	1.22
Communications Equipment	262	2,384	9.10	1,136	4.34
Motor Vehicles and Other Transportation Equipment[c]	152	809	5.32	1,791	11.78
Aircraft and Engines	48	501	10.44	653	13.60
Guided Missiles and Ordnance	16	173	10.81	103	6.44
Instruments	736	1,351	1.84	652	0.89
Total[d]	4,407	17,182	3.90	10,043	2.28

[a] Company *R&D* (1974) and patent (June 1976–March 1977) data are from Scherer (1983).
[b] Includes SIC 361, 362, 364, and 367.
[c] Includes SIC 371, 373, 374, 375, and 379.
[d] Includes only industries in this table.

tion between large-firm innovations and patents falls to 0.107, while that between small-firm innovations and patents is nearly twice as great. Similarly, the correlation between each of the *R&D* measures and small-firm innovations falls to less than 0.070 while that between each of the *R&D* measures and the large-firm innovations is not so greatly affected. In addition, the correlation between patents and each of the *R&D* measures falls to below 0.100 for the low-innovative industries. In the high-innovative industries, defined as those with at least fifteen innovations, the correlation between small-firm innovations and patents is also considerably weaker, as is that between the *R&D* measures and small-firm innovations.

Table 4 provides a more detailed comparison of an input measure—company *R&D* expenditures (from the FTC Line of Business Survey)—and two output measures, the total number of innovations and the total number of patented inventions between June 1976 and March 1977 (from Scherer, 1983). As the simple correlation of 0.74 indicates, there is a fairly strong relationship between the ratio of patents-per-innovation and *R&D* expenditures (millions of dollars)-per-innovation.[8] Of particular interest is that the patent-per-innovation ratio in drugs is about one-half of that in the rest of the broader chemical sector. This is remarkably similar to the pattern of patents-per-scientists and engineers identified by Robert E. Evenson (1984). The relatively high ratios of patents-per-innovation in the chemical and petroleum sectors, and the relatively low ratios in the computers, electrical machinery, lumber, and instruments sectors is explained by Edwin Mansfield (1984, p. 462): "The value

---

[8] Patents represent 59 percent of all invention patents issued to U.S. corporations during the sample period and 61 percent of patents issued to industrial corporations (Scherer, 1983).

TABLE 5—WEIGHTED MEAN NUMBER OF TOTAL INNOVATIONS, SMALL-FIRM INNOVATION, AND EMPLOYMENT SHARES BY TWO-DIGIT SECTOR[a]

Sector	Total Innovations	Small-Firm Innovation Share	Small-Firm Employment Share
Food	3.739	0.228	0.220
	(1.977)	(0.052)	(0.175)
Textiles	0.333	0.133	0.325
	(0.062)	(0.161)	(0.156)
Apparel	0.576	0.099	0.565
	(0.605)	(0.159)	(0.220)
Lumber	0.647	0.088	0.564
	(0.338)	(0.228)	(0.264)
Furniture	5.539	0.365	0.596
	(0.422)	(0.074)	(0.264)
Paper	3.588	0.161	0.234
	(0.103)	(0.022)	(0.199)
Printing	1.471	0.191	0.524
	(0.368)	(0.048)	(0.245)
Chemicals	17.929	0.313	0.135
	(0.014)	(0.041)	(0.163)
Petroleum	4.800	0.400	0.113
	(2.400)	(0.200)	(0.163)
Leather	0.546	0.273	0.422
	(0.776)	(0.230)	(0.182)
Stone,	2.185	0.219	0.347
Clay, and Glass	(0.036)	(0.013)	(0.264)
Primary Metals	2.846	0.276	0.227
	(1.631)	(0.043)	(0.167)
Fabricated Metal	9.556	0.388	0.524
Products	(4.301)	(0.038)	(0.192)
Machinery	25.886	0.485	0.289
(nonelectrical)	(1.355)	(0.014)	(0.199)
Electrical	20.946	0.411	0.164
Equipment	(1.824)	(0.008)	(0.155)
Transportation	9.000	0.149	0.075
Equipment	(1.000)	(0.029)	(0.325)
Instruments	56.615	0.506	0.221
	(16.344)	(0.017)	(0.121)

[a] The sector means are weighted averages (by 1977 value-of-shipments) of the four-digit SIC industry values. The small-firm innovation share is defined as the percentage of total innovations contributed by small firms. The small-firm employment share is defined as the percentage of industry employment accounted for by firms with fewer than 500 employees. The data come from the U.S. Small Business Administration, Small Business Data Base, 1982. Standard deviations are listed in parentheses.

and cost of individual patents vary enormously within and across industries... Many inventions are not patented. And in some industries, like electronics, there is considerable speculation that the patent system is being bypassed to a greater extent than in the past. Some types of technologies are more likely to be patented than others."

A broader comparison of the distribution of the innovations classified according to firm size, both between and within two-digit SIC sectors is presented in Table 5. For example, the first column shows that the mean number of total innovations in the 47 industries (weighted by the 1977 value-of-shipments) comprising the food sector was 3.74. The second column shows the small-firm innovation share, measured as the percentage of total innovations contributed by small firms, and indicates that in certain sectors the mean number of innovations contributed by the large firms far exceeded the mean number contributed by small firms, while in other sectors the pattern is reversed. As the third column indicates, in about half of the sectors the small-firm employment share exceeds the small-firm innovation share, and in the other half the innovation share of small firms exceeds the employment share.

II. The Empirical Model

We follow others (Jaffe, 1986; Scherer, 1982; and Mansfield, 1981) and assume innovative output is related to innovation-inducing inputs in the previous period according to a log relationship. The independent variables we expect to be important are: *R&D* expenditures (either total or company), 1977; capital intensity, measured as gross assets divided by value-of-shipments, 1977; the four-firm concentration ratio, 1977; the mean percentage of employees belonging to a union between 1973 and 1975; advertising expenditures divided by value-of-shipments; the share of 1977 industry employment accounted for by large firms (enterprises with more than 500 employees); a measure of skilled labor, defined as the percentage of employment consisting of professional and kindred workers, plus managers and administrators, plus craftsmen and kindred workers, 1970. Finally, to control for industry size, the 1977 value-of-shipments is included in the empirical model. All variable sources are listed in the Appendix.

Concentration, advertising, and capital intensity have all been hypothesized to facilitate appropriability and therefore encourage innovation (William S. Comanor, 1967; and Kamien and Schwartz, 1975). More recently, several models have been developed arguing that unions capture rents from intangible capital investments, and, in particular, those accruing from innovation-producing *R&D* (Robert A. Connolly, Barry Hirsch, and Mark Hirschey, 1986; and Hirsch and Albert N. Link, 1986). To the extent to which unions are successful in such rent-seeking activities, the ease of appropriability by the innovative firm is clearly reduced. Therefore, a negative relationship between unionization and total innovative activity is expected.

By substituting the number of large-firm innovations and small-firm innovations as dependent variables into the regression model, the hypothesis by Winter (1984) that two technological regimes exist can be examined. According to Winter (1984, p. 297), "An entrepreneurial regime is one that is favorable to innovative entry and unfavorable to innovative activity by established firms; a routinized regime is one in which the conditions are the other way round." To the extent to which innovations from small and large firms emanate from different technological regimes, a difference in the parameters in the small- and large-firm innovation equations would offer support for Winter's hypothesis that large- and small-firm innovations are promoted under differing economic and technological conditions.[9] The relative small-firm innovative advantage "is likely to be roughly proportional to the number of people exposed to the knowledge base from which innovative ideas might derive." (Winter, 1984, p. 297)

While the exact dynamic relationship between the endogenous and exogenous variables is not known, this should not pose a significant problem in the context of cross-sectional analysis where each of the industry-specific attributes remains relatively invariant over time. While this model assumes the traditional view that technical change is endogenous to market structure, Partha Dasgupta and Joseph Stiglitz (1980) argue that, in fact, certain aspects of market structure are endogenous to technical change.

III. Regression Results

Using the logged value of the number of total innovations in 1982 as the dependent

---

[9] While we have seen a major shift in the twentieth century, with the large firm and the industrial research laboratory playing a much more important role in innovation, individual entrepreneurs, who are not associated with established firms (Joseph Schumpeter, 1934, p. 62), continue to survive and thrive in a number of industries (Kenney Martin, 1986).

TABLE 6—REGRESSIONS OF LOG OF TOTAL NUMBER OF INNOVATIONS
(*t*-statistics in parentheses)

	1	2
Log Total *R&D*	0.364	–
	(6.434)[a]	
Log Company *R&D*	–	0.409
		(7.424)[a]
Log Capital/Output	0.007	−0.040
	(0.045)	(−0.278)
Log Concentration	−0.613	−0.631
	(−3.853)[a]	(−4.061)[a]
Log Unionization	−0.460	−0.425
	(−3.078)[a]	(−2.907)[a]
Log Advertising	−0.061	−0.056
	(−1.169)	(−1.096)
Log Large-Firm Employment Share	0.438	0.400
	(2.170)[a]	(2.023)[a]
Log Skilled Labor	0.505	0.559
	(2.257)[a]	(2.620)[a]
Log Industry Size	0.032	0.621
	(0.372)	(4.301)[a]
Constant	1.883	1.821
	(1.251)	(1.492)
Sample Size	247	247
R^2	0.457	0.482
F	25.040[a]	27.730[a]

[a] Statistically significant at the 95 percent level of confidence, two-tailed test.

variable, the cross-section regressions for 247 four-digit SIC manufacturing industries are shown in Table 6. While equation (1) uses the measure of total expenditures on *R&D*, equation (2) uses company expenditures on *R&D*. There are two major results from this table. First, the negative coefficient of concentration suggests that lower, and not higher, levels of concentration tend to be associated with increased innovation activity. While the emergence of a negative and significant relationship is consistent with the findings of Connolly and Hirschey (1984), it is a much stronger result than found by Scherer (1965). Second, the elasticity of both company and total *R&D* with respect to the total number of innovations is considerably less than one. Even when the measure of company *R&D*, which apparently explains innovative activity better than total *R&D*, is used in equation (2), the elasticity of *R&D* remains below 0.5. This is somewhat less than the elasticity of about one found between firm *R&D* and patents by Bound et al. (1984).

The positive coefficient of the large-firm employment share indicates that the greater the proportion of an industry consisting of firms with more than 500 employees, the greater is the innovation activity. Of the other variables representing appropriability, only unionization apparently is negatively associated with the total number of innovations, providing support for the Hirsch and Link (1986) hypothesis.[10]

Table 7 shows the results for analogous regressions where the log of the number of large-firm innovations, log of the number of small-firm innovations, and the small-firm innovation share are the dependent variables. Since the error terms in the large- and small-firm innovation regressions are presumably interrelated, the generalized least squares method for seemingly unrelated regressions was estimated using the Aitken procedure (Jan Kmenta, 1971). Skilled labor and advertising apparently have disparate relationships with small- and large-firm innovations. While skilled labor is positively and advertising negatively associated with small-firm innovations, neither is statistically associated with large-firm innovations. Although concentration is negatively associated with both large- and small-firm innovation, the elasticity of concentration with respect to small-firm innovations is more than double that for large firms. A perhaps somewhat surprising result is that not only is

---

[10] Regressions analogous to those in Table 6 were estimated substituting the patent measure used in Table 3 as the dependent variable. The results are quite similar to those using the innovation measure in Table 6. Total *R&D*, capital/output, the large-firm employment share, and skilled labor are found to be positively related to the number of patents, while concentration is negatively related to patents. Neither advertising, which has a negative coefficient, nor unionization, which has a positive coefficient, can be considered statistically significant. Thus, substituting the patent measure for the innovation measure yields very similar results. Only unionization and capital/output have different qualitative relationships between the patent and innovation measures.

TABLE 7—REGRESSIONS OF THE LOG OF THE NUMBER OF LARGE- AND SMALL-FIRM INNOVATIONS, AND THE SMALL-FIRM INNOVATION SHARE (*t*-statistics listed in parentheses)[a]

Equation	Total *R&D*	Company *R&D*	Concentration	Capital/ Output	Unionization	Advertising	Large-Firm Employment Share	Skilled Labor	Industry Size	Constant	Sample Size	R^2	F
1 Large-Firm Innovations	0.251	–	−0.276	0.107	−0.415	−0.042	0.371	0.294	0.150	0.020	247	0.372	–
	(4.711)^{c}		(−1.836)^{b}	(0.763)	(−2.937)^{c}	(−0.854)	(1.943)^{b}	(1.388)	(1.825)^{b}	(0.017)			
2 Small-Firm Innovations	0.278	–	−0.705	−0.060	−0.442	−0.082	0.437	0.383	−0.158	3.802	247	0.366	–
	(5.850)^{c}		(2.036)^{c}	(−0.481)	(−3.517)^{c}	(−1.866)^{b}	(2.572)^{c}	(2.036)^{c}	(−2.155)^{c}	(1.056)			
3 Large-Firm Innovations	–	0.294	−0.294	0.078	−0.388	−0.038	0.339	0.318	0.110	−0.022	247	0.393	–
		(5.612)^{c}	(−1.987)^{c}	(0.568)	(−2.785)^{c}	(−0.787)	(1.802)^{b}	(1.567)	(1.361)	(−0.019)			
4 Small-Firm Innovations	–	0.324	−0.723	−0.093	−0.412	−0.078	0.401	0.411	−0.202	3.756	247	0.398	–
		(7.015)^{c}	(−5.556)^{c}	(−0.769)	(−3.356)^{c}	(−1.813)^{b}	(2.423)^{c}	(2.300)^{c}	(−2.842)^{c}	(3.669)^{c}			
5 Small-Firm Innovation Share	−0.389	–	−0.004	−0.505	−0.001	−7.306	0.355	0.785	−0.154	0.168	247	0.117	3.94^{b}
	(−1.765)^{b}		(−2.811)^{c}	(−3.077)^{c}	(0.569)	(−0.456)	(2.535)^{c}	(2.881)^{c}	(−0.367)	(1.459)			
6 Small-Firm Innovation Share	–	−0.138	−0.004	−0.484	−0.001	−7.194	0.330	0.666	−0.400	0.209	247	0.116	3.92^{b}
		(−0.463)	(−2.837)^{c}	(−2.934)^{b}	(0.637)	(−0.446)	(2.327)^{c}	(2.482)^{c}	(−0.903)	(1.834)^{b}			

[a] The method of generalized least squares (Aitken procedure) was used to estimate large-firm innovations and small-firm innovations. The logs for all explanatory variables are used to estimate log large-firm innovations and log small-firm innovations but not for the small-firm innovation share.

[b] Statistically significant at the 90 percent level of confidence, two-tailed test.

[c] Statistically significant at the 95 percent level of confidence, two-tailed test.

the coefficient of the large-firm employment share positive and significant for small-firm innovations, but it is actually greater in magnitude than for large firms. This suggests that, *ceteris paribus*, the greater extent to which an industry is composed of large firms, the greater will be the innovative activity, but that increased innovative activity will tend to emanate more from the small firms than from the large firms. Perhaps this indicates that, in industries composed predominately of large firms, the existing small firms must resort to a strategy of innovation in order to remain viable. This is consistent with the findings of Richard E. Caves and Thomas A. Pugel (1980) that smaller firms in an industry tend to perform better if they use different strategies from those followed by the larger firms.[11]

Equations (5) and (6) address a slightly different question. Rather than considering what determines the extent of innovative activity in small firms, small-firm innovative activity is examined relative to that in the entire industry.[12] The small-firm innovation share, or the number of small-firm innovations divided by the number of total innovations, is used as the dependent variable. The negative coefficients of total *R&D*, concentration, and capital/output, suggest that the number of small-firm innovations tends to be small relative to that in the entire industry in industries which are high in *R&D*, capital-intensive, and which are concentrated. The positive coefficients of the large-firm employment share and skilled labor imply that the innovation activity of small firms is high relative to the industry level in in-

---

[11] The disparate effect of the large-firm employment share on small- and large-firm innovation rates is perhaps implied by Scherer (1980, p. 422), "...the most favorable industrial environment for rapid technological progress would appear to be a firm size distribution that includes a predominance of companies with sales below $500 million, pressed on one side by a horde of small, technologically oriented enterprises bubbling over with bright new ideas and on the other by a few larger corporations with the capacity to undertake exceptionally ambitious developments."

[12] For further analysis of the difference between large- and small-firm innovation rates, see our 1987 paper.

TABLE 8—HYPOTHESES TESTS FOR THE DIFFERENCES IN THE DETERMINANTS OF INNOVATION BETWEEN LARGE AND SMALL FIRMS (BASED ON REGRESSIONS 3 AND 4 IN TABLE 7)

Variable Hypothesized[a]	SSR[b]	F
Company *R&D*	118.21	3.665[d]
Capital/Output	120.80	0.279
Concentration	118.25	3.625[d]
Unionization	120.71	0.398
Advertising	120.73	0.359
Large-Firm Employment Share	120.73	0.359
Skilled Labor	118.23	3.630[d]
All Variables[c]	104.276	2.731[d]

[a] The variable hypothesized and therefore constrained to have an identical effect on large-firm innovations and small-firm innovations. Each hypothesis is tested by comparing the sum of the squares explained by the regression from the regression with the restricted coefficient to that from the unrestricted regression. For further explanation of the hypothesis test, see Kmenta (1971).

[b] SSR refers to the sum of squares explained by the estimated regressions for small-firm innovations with the restricted coefficient.

[c] The coefficients of all the explanatory variables are constrained to be identical for the large-firm innovation and small-firm innovation equations.

[d] Statistically significant at the 95 percent level of confidence.

dustries in which there is only a small share of small firms, and in which skilled labor plays an important role.

The results from Table 7 do not support an unequivocal conclusion regarding the exact differences in innovation behavior between large and small firms. While the first four equations imply that skilled labor, advertising, and perhaps concentration and the large-firm employment share, account for the different patterns in innovative activity between large and small firms, the last two equations suggest that total *R&D*, concentration, capital/output, large-firm employment share, and skilled labor account for the differences in relative innovative activity. Thus, in Table 8 the hypotheses that each variable has an identical effect on small- and large-firm innovation activity are tested. The resulting *F*-test is statistically significant for skilled labor, company *R&D*, and concentration, as well as for the hypothesis that all of the variables jointly have an identical effect on small- and large-firm innovation activity. Thus, while there still remains ambiguity regarding the exact patterns between *R&D*, appropriability, and innovation activity for large and small firms, there is considerable support for Winter's (1984) hypothesis that different economic and technological regimes may account for at least some of the differences between the innovation activity of large and small firms.

IV. Conclusion

We have introduced a new and more direct measure of innovative activity in this paper and found that, while the total number of innovations is closely related to *R&D* expenditures (particularly after excluding federal *R&D*), and patented inventions, especially at the more aggregated level, the exact relationship between *R&D* and innovation is somewhat different from that between *R&D* and patents. While Bound et al. (1984) found that, based on firm data, the elasticity of *R&D* expenditures with respect to patented inventions is about one, we find that the number of innovations increases with increased industry *R&D* expenditures but at a decreasing rate. Similarly, while the literature has found a somewhat ambiguous relationship between concentration and various measures of technical change, our results are unequivocal—industry innovation tends to decrease as the level of concentration rises. In addition, there is evidence supporting the hypothesis that unionization is negatively related to innovation activity. Of course, these results are at best an introduction of the innovation data and need to be qualified, particularly in light of the likely simultaneity between industry technology and market structure. Future research may be able to incorporate the endogeneity of market structure in a manner suggested by Levin and Peter Reiss (1984).

This paper has also provided evidence supporting Winter's (1984) prediction that the innovation activity of small and large firms responds to considerably different technological and economic environments.

These findings are not without ambiguity. Further research should apply the patent measure in a similar framework to determine whether its performance differs from that of the innovation measure used here.

APPENDIX: DATA SOURCES AND FURTHER EXPLANATIONS

The data on innovations and employment for small and large firms come from the U.S. Small Business Administration (Edwards and Gordon, 1984). These data can be obtained upon request from the authors. The U.S. Department of Commerce, Bureau of the Census, *Annual Survey of Manufactures*, 1977, *Industry Profiles*, Washington: USGPO, issued 1981, is the source for capital/output (gross assets divided by value-of-shipments), concentration, and value-of-shipments. The advertising measure was derived by using the value-of-shipments data described above and advertising expenditures, from the 1972 U.S. Input-Output Table.

The unionization data are from Richard B. Freeman and James L. Medoff (1979). The percentage of total employment that is unionized for three-digit SIC industries between 1973 and 1975 is reported. We repeat these three-digit SIC values at the four-digit level. The measure of skilled labor is from the U.S. Department of Commerce, Bureau of the Census, Census Population 1970, Subject Report PC (2)-7C, *Occupation by Industry*, Washington: USGPO, issued 1972.

The total *R&D* and company *R&D* data are from the 1977 Federal Trade Commision Line of Business Report. The patent measure used for the disaggregated three-digit SIC correlations in Table 3 is from the Office of Technology Assessment and Forecast of the U.S. Patent Office. Both the patent and skill measures are taken from the *U.S. International Trade Commission's Industrial Characteristics and Trade Performance Data Bank*, June 1975, Office of Economic Research, Washington, D.C. The USITC data bank is described by Edward J. Ray (1981) and these measures are available from the authors on request.

REFERENCES

Acs, Zoltan J. and Audretsch, David B., "Innovation, Market Structure, and Firm Size," *Review of Economics and Statistics*, November 1987, *69*, 567–75.

Bound, John, Cummins, Clint, Griliches, Zvi, Hall, Bronwyn H. and Jaffe, Adam, "Who Does *R&D* and Who Patents?," in Zvi Griliches, ed., *R&D, Patents, and Productivity*, Chicago: University of Chicago, 1984, 21–54.

Caves, Richard E. and Pugel, Thomas A., *Intraindustry Differences in Conduct and Peformance: Viable Strategies in U.S. Manufacturing Industries*, New York University Press, 1980.

Comanor, William S., "Market Structure, Product Differentiation, and Industrial Research," *Quarterly Journal of Economics*, November 1967, *81*, 639–57.

Connolly, Robert A., Hirsch, Barry T. and Hirschey, Mark, "Union Rent Seeking, Intangible Capital, and the Market Value of the Firm," *Review of Economics and Statistics*, November 1986, *68*, 567–77.

______ **and Hirschey, Mark**, "*R&D*, Market Structure and Profits: A Value-Based Approach," *Review of Economics and Statistics*, November 1984, *66*, 682–86.

Dasgupta, Partha and Stiglitz, Joseph, "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, June 1980, *90*, 266–93.

Edwards, Keith L. and Gordon, Theodore, J., "Characterization of Innovations Introduced on the U.S. Market in 1982," The Futures Group, U.S. Small Business Administration, Contract No. SBA-6050-0A-82, March 1984.

Evenson, Robert E., "International Invention: Implications for Technology Market Analysis," in Zvi Griliches, ed., *R&D, Patents, and Productivity*, Chicago: University of Chicago, 1984, 55–70.

Freeman, Richard B. and Medoff, James L., "New Estimates of Private Sector Unionism in the United States," *Industrial and Labor Relations Review*, January 1979, *32*, 143–74.

Griliches, Zvi, "Productivity, *R&D*, and Basic Research at the Firm Level in the 1970's," *American Economic Review*, March 1986, *76*, 141–54.

Hall, Bronwyn H., Griliches, Zvi and Hausman, Jerry A., "Patents and R&D: Is There a Lag?," *International Economic Review*, June 1986, *27*, 265–302.

Hirsch, Barry T. and Link, Albert N., "Labor Union Effects on Innovative Activity," unpublished manuscript, University of North Carolina-Greensboro, 1986.

Jaffe, Adam B., "Technological Opportunity and Spillovers of *R&D*: Evidence from Firms' Patents, Profits, and Market Value," *American Economic Review*, December 1986, *76*, 984–1001.

Kamien, Morton I. and Schwartz, Nancy L., "Market Structure and Innovation: A Survey," *Journal of Economic Literature*, March 1975, *13*, 1–37.

Kmenta, Jan, *Elements of Econometrics*, New York: Macmillan, 1971.

Kuznets, Simon, "Inventive Activity: Problems of Definition and Measurement," in R. R. Nelson, ed., *The Rate of Direction of Inventive Activity*, Princeton: Princeton University Press, 1962, 19–43.

Levin, Richard C., Cohen, Welsley M. and Mowery, David C., "*R&D* Appropriability, Opportunity, and Market Structure: New Evidence on Some Schumpeterian Hypotheses," *American Economic Review*, May 1985, *75*, 20–24.

______, **and Reiss, Peter C.,** "Tests of a Schumpeterian Model of *R&D* and Market Structure," in Zvi Griliches, ed., *R&D*, *Patents, and Productivity*, Chicago: University of Chicago Press, 1984, 462–64.

Mansfield, Edwin, "Comment on Using Linked Patent and *R&D* Data to Measure Interindustry Technology Flows," in Zvi Griliches, ed., *R&D*, *Patents, and Productivity*, Chicago: University of Chicago, 1984, 175–204.

______, "Composition of R and D Expenditures: Relationship to Size of Firm, Concentration, and Innovative Output," *Review of Economics and Statistics*, November 1981, *63*, 610–15.

Martin, Kenney, "Schumpeterian Innovation and Entrepreneurs in Capitalism," *Research Policy*, February 1986, *1*, 21–33.

Pakes, Ariel, "On Patents, *R&D*, and the Stock Market Rate of Return," *Journal of Political Economy*, April 1985, *93*, 390–409.

______, **and Griliches, Zvi,** "Patents and *R&D* at the Firm Level: A First Report," *Economics Letters*, 1980, *4*, 377–81.

______, **and Griliches, Zvi,** "Patents and *R&D* at the Firm Level: A First Look," in Zvi Griliches, ed., *R&D*, *Patents, and Productivity*, Chicago: University of Chicago, 1984, 55–70.

______, **and Schankerman, Mark,** "The Rate of Obsolescence of Patents, Research Gestation Lags, and the Private Rate of Return to Research Resources," in Zvi Griliches, ed., *R&D*, *Patents, and Productivity*, Chicago: University of Chicago, 1984, 73–88.

Ray, Edward J., "The Determinants of Tariffs and Nontariff Trade Restrictions in the United States," *Journal of Political Economy*, February 1981, *89*, 105–21.

Scherer, F. M., "The Propensity to Patent," *International Journal of Industrial Organization*, 1983, *1*, 107–28.

______, "Inter-Industry Technology Flows and Productivity Growth," *Review of Economics and Statistics*, December 1982, *64*, 627–34.

______, *Industrial Market Structure and Economic Performance*, Chicago: Rand McNally College Publishing, 1980.

______, "Market Structure and the Employment of Scientists and Engineers," *American Economic Review*, June 1967, *57*, 524–30.

______, "Firm Size, Market Structure, Opportunity, and the Output of Patented Inventions," *American Economic Review*, December 1965, *55*, 1097–125.

Schumpeter, Joseph, *The Theory of Economic Development*, Cambridge: Harvard University Press, 1934.

Winter, Sidney G., "Schumpeterian Competition in Alternative Technological Regimes," *Journal of Economic Behavior and Organization*, September-December 1984, *5*, 287–320.

Cash Versus Kind, Self-Selection, and Efficient Transfers

By CHARLES BLACKORBY AND DAVID DONALDSON*

This paper investigates second-best (transfers in kind) and third-best (subsidies and taxes) Pareto optima in a simple model where government lacks full information about consumer types (who is able, who is infirm). These Pareto optima rely on self-selection. We show that every third-best optimum other than the equal-income Walrasian equilibrium is Pareto-dominated by a second-best optimum and that standard "willingness-to-pay" criteria are inappropriate in this environment.

The second theorem of welfare economics guarantees that, in convex economies, Pareto-efficient allocations can be decentralized as Walrasian equilibria. Lump-sum transfers are required along with information about preferences and needs (consumption sets). Transfers of purchasing power are "efficient" in the sense that they move the economy from one Pareto optimum to another. This property is not always shared by transfers of goods and services when resale is difficult or impossible.

If information about preferences is not publicly available to the government, the superiority of transfers of purchasing power over transfers of goods and services disappears (see the discussions by Peter Diamond and James Mirrlees, 1978; Ronald Dye and Rick Antle, 1986; Milton Harris and Robert Townsend, 1981; Peter Hammond, 1987; Albert Nichols and Richard Zeckhauser, 1982; and Kevin Roberts, 1984). The reason is that people may attempt to claim transfers intended for others. For example, if rape victims are offered transfers of cash, everyone has an interest in qualifying. But if counseling is provided instead, only the intended beneficiaries will be interested in the transfer. As long as it is impossible or costly to distinguish victims from other people, there is an advantage to supplying the counseling service. This observation applies to many other social services, and to medical care and education.[1] In all these cases, resale of the service is not possible.

In this paper we present a very simple model that captures this phenomenon. A population consists of two types of individuals. We consider one person of each type. *Able* is a healthy person, and *Infirm* is ill.[2] There are only two commodities in the economy, yams and medical care. Both Able and

---

*Department of Economics, University of British Columbia, #997-1873 East Mall, Vancouver, B.C., Canada, V6T 1W5. An earlier version of this paper was prepared for the conference *Economic Models and Distributive Justice*, Brussels and Namur, January 8–10, 1987; we wish to thank the organizers, Louis Gevers and Claude d'Aspremont, and the participants, particularly Peter Hammond and our colleague John Weymark. Earlier versions were presented at Dalhousie University, the *Twelfth Canadian Economic Theory* meeting at Montreal, and U.B.C, and we thank the participants of these seminars. In addition we have benefited from detailed comments by Bill Schworm and two referees of this journal for which we are grateful. We thank Francine Lafontaine and Robert Picard for programming assistance, and the Social Sciences and Humanities Research Council of Canada for research support.

[1] Medical care in the real world is a complicated case. Physicians often act as their patients' agents, but are imperfectly informed about patients' health status (they typically know more about some things and less about others) and face incentives that may lead them to overuse some treatments. We abstract from these considerations in this paper. Education fits our model when individuals know more about their abilities and drives than government does. Public education usually is offered in kind and it is rationed differently to people of different proven abilities. Voucher schemes are a form of transfer in kind.

[2] Able and Infirm appeared, originally, in Cohen (1986). The model can be extended to arbitrary numbers of each type as long as different individuals of a given type receive equal treatment.

Infirm want yams; Infirm alone wants (needs) medical care. Medical care cannot be resold.

First-best Pareto optima, in this economy, require (as long as Infirm's consumption of both goods is positive) that Infirm's marginal rate of substitution be equal to the marginal rate of transformation. All efficient outcomes can be decentralized as Walrasian equilibria.[3] This result requires that full information about preferences be public knowledge.

It is possible that the government may not know which individuals are infirm and which able, but is limited instead to knowledge of the proportion of each type in the population. In that case, it may choose to offer two different consumption bundles, one with yams only, $(y_A, 0)$, and another with yams and medical care, (y_I, z), and allow individuals to choose between them. In the absence of moral qualms, Able will pose as Infirm if $y_I > y_A$ (we assume free disposal of medical care; Able need not accept treatment). Similarly, Infirm will pose as Able if the utility of $(y_A, 0)$ exceeds the utility of (y_I, z). *Second-best Pareto-optima* are Pareto-optima in this environment; efficiency considerations alone require that Able and Infirm select the bundles that are intended for them (see Section II below). Thus, *second-best optima* satisfy self-selection constraints. We show that *i*) there are some first-best optima that are second-best as well; *ii*) there are some second-best optima that are not first-best optima; *iii*) in all second-best optima that are not first-best, medical care is "overprovided" —the marginal rate of substitution is less than the marginal rate of transformation— and the quantity exceeds the first-best level; *iv*) standard "willingness-to-pay" criteria are inappropriate in this model.

All second-best optima are "envy-free" in the sense that neither Able nor Infirm would prefer to consume the other's bundle; however, these optima require *rationing*.[4] Infirm may not control the amount of medical care he or she receives.

It is possible to describe *third-best Pareto optima*, allocations where medical care can be bought and sold at a government-set market price, and where information remains imperfect. Self-selection now requires that the values of the two consumption bundles (incomes) should be the *same* (if not, one will pose as the other). This is a stronger restriction than freedom from envy and requires that there is no "income-envy." The price of medical care may be used to affect the distribution of welfare, and a set of third-best optima can be found. We show that *v*) there is only one third-best allocation that is first-best—the equal-income Walrasian equilibrium; *vi*) no second-best optimum other than this one is a third-best optimum; *vii*) each third-best optimum other than the equal-income Walrasian equilibrium is Pareto-dominated by a second-best optimum.

These results suggest that, in these circumstances, market solutions are inferior to rationing solutions even if subsidization or taxation of medical care is allowed.

Our analysis is welfarist throughout—only utilities, and not their sources, count in the process of social evaluation. At the same time, it is clear that the Pareto optima chosen by standard welfarist criteria, such as utilitarianism or maximum utility, may appear to have nonwelfarist characteristics. Second-best optima may appear to "force" Infirm to overconsume medical care (he or she would purchase less if faced with the Walrasian price and an equivalent income). We believe that it is important to realize that our intuitions about social policy are formed in a world of imperfections. Thus, accusations of paternalism against those who favor rationed social provision of goods like medical care may be misplaced.

I. First-Best Pareto Optima

$y_A \geq 0$ is Able's consumption of yams and $(y_I, z) \geq (0,0)$ is Infirm's consumption of yams and medical care, respectively. Able's preferences are represented by the utility function

---

[3] All agents are price takers.

[4] By rationing, we mean that the quantity actually consumed must be less than or equal to the fixed quantity. This occurs in the model because of our assumption of free disposal.

$$(1)\qquad u_A = y_A,$$

where u_A is Able's (ordinal) utility.[5] Infirm's preferences are represented by the utility function U whose image is

$$(2)\qquad u_I = U(y_I, z).$$

We assume that U is increasing, twice continuously differentiable, and strongly quasi concave.[6] In addition, we assume that y and z are normal goods for Infirm. The economy is constrained by a production possibility frontier given by

$$(3)\qquad y_A + y_I + z = k.^{7}$$

The set of first-best allocations and the first-best utility possibility frontier (FBF) may be characterized by maximizing

$$(4)\qquad U(y_I, z),$$

subject to the production possibility frontier, (3), and to

$$(5)\qquad y_A = u_A, \qquad 0 \le u_A \le k.$$

The solution is given by

$$(6)\qquad u_I^f = F(u_A), \qquad 0 \le u_A \le k.$$

F is clearly a decreasing function; (6) describes the first-best frontier (the FBF).

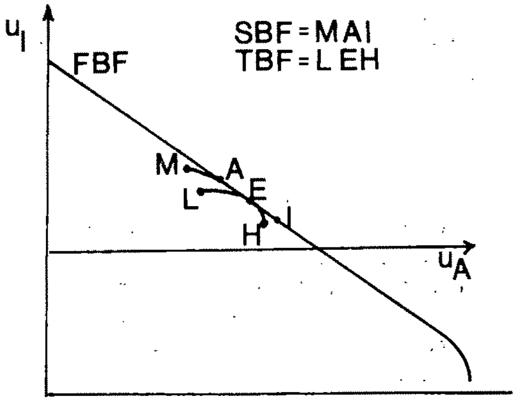


FIGURE 1

As long as full information about preferences (who is able and who infirm) is public knowledge, first-best allocations can be decentralized as Walrasian equilibria, with the price of medical care relative to yams equal to 1, the marginal rate of transformation. If (y_A^f, y_I^f, z^f) is a first-best allocation, then Able and Infirm must have purchasing powers m_A^f and m_I^f (y is chosen as numeraire and its price set equal to 1), where

$$(7)\qquad m_A^f = y_A^f$$

and

$$(8)\qquad m_I^f = y_I^f + z^f.$$

If the values of their endowments (there is no profit because of constant returns) differ from m_A^f and m_I^f, a lump-sum transfer must be made from one to the other.

As an example, we consider the utility function for Infirm[8]

$$(9)\qquad U(y_I, z) = y_I - e^{1-z},$$

and set $k = 6$. The FBF is shown in Figure 1. Along the frontier, $z^f = 1$ as long

---

[5] Because preferences are ordinal and Able enjoys only one commodity, there is no loss of generality in choosing this representation.

[6] See W.E. Diewert, M. Avriel, and I. Zang (1981). U is strongly concave if and only if, for all

$$(y_I, z) \in E_+^2, \qquad (v_1, v_2) \neq (0,0),$$

$$(y_I + v_1, z + v_2) \in E_+^2, \qquad v_1 U_y + v_2 U_z = 0$$

implies that $v_1^2 U_{yy} + 2 v_1 v_2 U_{yz} + v_2^2 U_{zz} < 0$.

[7] This linear production possibility frontier makes analysis of the problem simpler and does not entail any loss of generality with respect to those aspects of the problem in which we are interested.

[8] The utility function in (9) does not make y and z normal goods at all income levels (instead they are noninferior). This does not complicate the example unduly. Normality of goods simplifies the analysis considerably in the general case; its main function is to ensure that y_I and z are positive along the FBF as long as $u_A < k$.

as $0 \le u_A \le 5$, and $u_I^f = 4 - u_A$. For $u_A > 5$, $y_I^f = 0$ and $z^f = 6 - u_A$ so that $u_I^f = -\exp[1-(6-u_A)]$.

II. Second-Best Pareto Optima

If the government cannot tell who is Able and who is Infirm, then a problem arises. If Able and Infirm are selfish and unprincipled, each will pose as the other if he or she gains by it. Consequently, the government must take account of self-selection when allocating resources.

We suppose that the government offers two bundles $(y_A, 0)$ and (y_I, z) to the two people. For self-selection, Able must be allocated sufficient yams to prevent him or her from pretending to be infirm. That is,

$$y_A \ge y_I. \tag{10}$$

If Able receives *many* more yams than Infirm, it is possible that Infirm will choose to forego the offered medical care. Thus, self-selection requires

$$U(y_I, z) \ge U(y_A, 0). \tag{11}$$

In our example, (11) becomes

$$y_I - e^{1-z} \ge y_A - e. \tag{12}$$

The constraints are illustrated in Figure 2 for a fixed value of y_A; the two constraints cannot bind simultaneously unless $z = 0$.

Second-best Pareto optima are the Pareto optima for this rationing scheme in the above restricted information environment. All second-best Pareto optima must satisfy the self-selection constraints (10) and (11). To see this, suppose that an allocation has both people pretending to be able and consuming $(y, 0)$. If $z > 0$, this is inefficient and hence not a second-best optimum; if $z = 0$, the constraints are satisfied. Suppose, on the other hand, that both people pretend to be infirm, and are offered (y, z). This is equivalent, in utilities, to the bundles $(y, 0)$ for Able and (y, z) for Infirm which satisfy the constraints. If $z > 0$, resources are saved by the change.

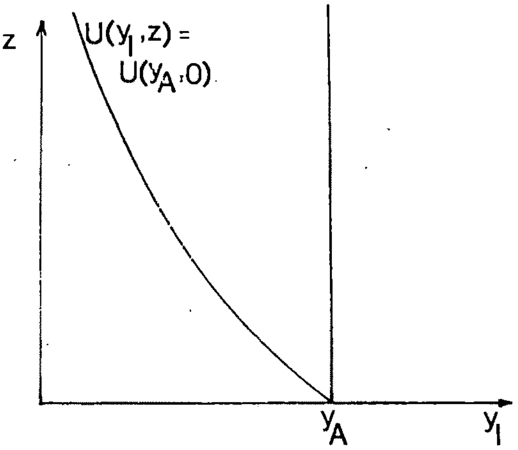


FIGURE 2

First we demonstrate that some points on the FBF are also on the second-best frontier (SBF). Using (5), rewrite the self-selection constraint (11) as

$$u_I \ge U(u_A, 0). \tag{13}$$

At $u_A = k$, $u_I^f = F(k) = U(0,0)$ so that (13) is not satisfied. Define the largest value of u_A which satisfies (13) by

$$\bar{u}_I^f = F(\bar{u}_A) = U(\bar{u}_A, 0). \tag{14}$$

Since F is decreasing, the constraint (13) is satisfied at all points on the FBF with $u_A \le \bar{u}_A$ and is not satisfied when $u_A > \bar{u}_A$. Next rewrite the self-selection constraint (10) for points on the FBF by

$$u_A \ge y_I^f, \tag{15}$$

where y_I^f is the first-best value of y_I corresponding to u_A. Since y is a normal good for Infirm, its consumption is strictly decreasing along the FBF as u_A increases. This allows us to define the smallest value of u_A which satisfies the self-selection constraint (10) as

$$\underline{u}_A = y_I^f. \tag{16}$$

Since y is a normal good for Infirm, the solution to (16) is unique. Clearly, the self-selection constraint (15) is satisfied for all

$u_A \geq \underline{u}_A$.[9] It is shown in the Appendix that $\underline{u}_A < \bar{u}_A$. Hence all first-best optima with $\underline{u}_A \leq u_A \leq \bar{u}_A$ are also second-best Pareto optima.

In our example, $\bar{u}_A$ is $(2+(e/2)) = 3.4$, and $\underline{u}_A$ is 2.5. All points on the FBF between A and I in Figure 1 are second-best optima.

There are second-best Pareto optima outside this range, and they are different from the corresponding first-best optima since one of the self-selection constraints binds. Each second-best Pareto optimum must maximize

$$u_I = U(y_I, z) \tag{17}$$

subject to (3), (5), and the self-selection constraints (10) and (11). For $u_A = y_A > \bar{u}_A$, no solution to this problem exists since (11) cannot be satisfied. For values of u_A such that $0 \leq u_A \leq \bar{u}_A$, we write the solution to this problem as

$$u_I^s = S(u_A). \tag{18}$$

All second-best Pareto optima satisfy (18), but not all utility pairs satisfying (18) are second-best Pareto optima. We define u_A^m as the value of u_A that maximizes u_I^s. Then the second-best frontier is given by (18) for $u_A^m \leq u_A \leq \bar{u}_A$.[10]

THEOREM 1: *(i) There are no second-best optima with* $u_A > \bar{u}_A$; *(ii) For all* $u_A, \underline{u} \leq u_A \leq \bar{u}_A$, *the second-best and first-best optima coincide; (iii) There are the second-best optima with* $u_A < \underline{u}_A$; *in all of these allocations*

$$MRS_{yz} = \frac{U_z(y_I, z)}{U_y(y_I, z)} < 1, \tag{19}$$

and the second-best level of z exceeds the first best.

PROOF: See the Appendix.

The proof of the theorem shows that $u_A^m < \underline{u}_A$ and that S is decreasing for all $u_A \geq u_A^m$. In the range $u_A^m \leq u_A < \underline{u}_A$, the second-best allocation for Infirm is $(u_A, k - 2u_A)$. The reason for this is that the first-best level of $y_I, y_I^f > u_A = y_A$, which is ruled out by self-selection. Hence, in this part of the SBF,

$$u_I^s = S(u_A) = U(u_A, k - 2u_A). \tag{20}$$

If $u_A^m > 0$, then the first-order condition for a maximum holds and is given by

$$U_y(u_A^m, k - 2u_A^m) - 2U_z(u_A^m, k - 2u_A^m) = 0. \tag{21}$$

Here the $MRS_{yz} = 1/2$ so that Infirm's shadow price (marginal willingness-to-pay for medical care) is half of marginal cost. Further reductions of y_I with corresponding increases in z reduce Infirm's utility along with Able's, so there are no second-best optima with $u_A < u_A^m$.

If levels of utility are interpersonally comparable, then the maximin utility criterion may be applied. If

$$u_I^m = U(u_A^m, k - 2u_A^m) \leq u_A^m, \tag{22}$$

then this point on the SBF will be selected.

In the example (see Figure 1), for $u_A < 5/2$, the second-best frontier is given by

$$u_I = u_A - e^{(2u_A - 5)}, \tag{23}$$

and u_I reaches a maximum at

$$u_A^m = \frac{5 - \ln 2}{2} = 2.2, \tag{24}$$

where $u_I^m = 1.7$. The second-best frontier is given by (23), $2.2 \leq u_A \leq 2.5$ and coincides with the FBF for $2.5 \leq u_A \leq 3.4$. Below $u_A = 2.5$,

$$MRS_{yz} = e^{(2u_A - 5)} < 1. \tag{25}$$

---

[9] If we dispense with the normality assumption, there may be several values of $u_A = y_I$ satisfying (16). The "true" second-best regions may therefore not be connected.

[10] It is straightforward to show that strong quasi-concavity of U implies strong quasi-concavity of u_I^s in u_A. It follows that there is exactly one u_A^m satisfying (18), and that all combinations (u_A, u_I^s) with $u_A^m \leq u_A \leq \bar{u}_A$ are second-best optima. If $u_A^m = 0$, then all $(u_A, u_I^s), 0 \leq u_A \leq \bar{u}_A$ are second-best optimal.

The second-best frontier (SBF) is drawn in Figure 1 and is MAI.

This second-best result has important consequences for the provision of social services. A standard cost-benefit evaluation of in-kind programs is a marginal willingness-to-pay test. That is, one asks how many yams the recipient would give up to obtain a unit increase in z.[11] Locally this is given by the marginal rate of substitution between y and z. Since the $MRS_{yz} < 1$ at second-best optima where the self-selection constraint (10) binds, this means that willingness-to-pay or marginal benefit tests for optimal provision are inappropriate.

III. Third-Best Pareto Optima

In second-best optima, consumers are rationed. They must choose a fixed quantity of medical care and there is no way for them to sell it for yams or to buy more. In this section we investigate the case of market solutions in the second-best information environment.

First-best optima allow Walrasian markets to be used (the second theorem of welfare economics) with incomes (after lump-sum transfers) given by (7) and (8). If, however, one agent's money income is greater than the other's then the low-income person will have an incentive to lie about his or her health status. Consequently, an incentive problem exists with all first-best market solutions other than the equal-income Walrasian equilibrium.

Third-best Pareto optima can be characterized as follows. The government sets a price q for medical care (the price of yams is set equal to one) and an income m which both Able and Infirm enjoy. Because both Able and Infirm have the same budget set, neither will be tempted to pose as the other (there is no "income envy"). Able will spend all of his or her income on yams. Infirm will choose $\tilde{y}_I = Y(q,m)$ and $\tilde{z}_I = Z(q,m)$ to maximize $U(y_I, z)$ subject to $y_I + qz \leq m$. Infirm's maximized utility, given q and m, is represented by the indirect utility function

$$(26)\quad \begin{aligned} u_I &= V(q,m) \\ &:= U(Y(q,m), Z(q,m)) \\ &= \max_{y_I, z}\{U(y_I, z) \mid y_I + qz \leq m\}. \end{aligned}$$

All third-best Pareto optima must solve the problem of choosing (q,m) to maximize

$$(27)\qquad V(q,m)$$

subject to

$$(28)\qquad y_A = u_A = m,\ 0 \leq u_A \leq k,$$

and

$$(29)\qquad y_A + Y(q,m) + Z(q,m) \leq k.$$

Solutions will exist for some but not all u_A in $[0,k]$. For all u_A where a solution exists, we write

$$(30)\qquad u_I^t = T(u_A).$$

By restricting the domain of T to the region where T has a nonpositive slope, the third-best frontier is found.

THEOREM 2: (*i*) *The equal-income Walrasian equilibrium is a third-best Pareto optimum and it is the only first-best optimum that is also a third-best optimum*; (*ii*) *No second-best optimum other than the equal-income Walrasian equilibrium is a third-best optimum*; (*iii*) *There are third-best optima other than the equal-income Walrasian equilibrium*; (*iv*) *Every third-best optimum other than the equal-income Walrasian equilibrium is Pareto dominated by a second-best optimum.*

PROOF: See the Appendix.

The proof of Theorem 2: (*iii*) shows that there are third-best optima on both sides of the equal-income Walrasian equilibrium (EIWE). If (u_A^E, u_I^E) are the utilities at the EIWE, then if $u_A < u_A^E$, a third-best optimum must have $V(\tilde{q}, u_A) > V(1, u_A^E) = u_I^E$, where $\tilde{q}$ is the optimal choice of q. In this

---

[11] See, for example, the argument by Maria Schmundt, Eugene Smolensky, and Leanna Stiefel (1975).

TABLE 1

Type of Pareto Optimum	Lowest u_A	Highest u_A	Lowest u_I	Highest u_I
First-Best	0	6	−2.8	4
Second-Best	2.1	3.4	.6	1.7
Third-Best	2.3	3.2	.6	1.3

case, therefore, $\tilde{q} < 1$ and *medical care is subsidized*. Third-best optima that favor Infirm relative to the EIWE provide low-cost medical care. A similar argument establishes that $\tilde{q} > 1$ for all third-best optima with $u_A > u_A^E$. Third-best optima that favor Able relative to the EIWE *tax medical care*.

Theorem 2 makes it clear that, when there are information constraints, markets may be less attractive than rationing. Rationing permits a whole range of first-best equilibria to survive as second-best. In addition, all third-best optima other than the equal-income ones are Pareto inferior to some second-best optimum.

This result is illustrated by our example. In Figure 1, the third-best frontier (TBF = LEH) lies inside the second-best, touching it at E, the equal-income Walrasian equilibrium. The example is summarized in Table 1.

The third-best maximin utility for Infirm is $u_I = 1.3$. This is a loss of maximin utility over second-best ($u_I^m = 1.7$) or first-best ($u_I = u_A = 2.0$).

IV. *Ex Ante* and *Ex Post* Optimality

The discussion above is set in a certain world. We can imagine, however, a certain outcome *ex post*—one person able, one infirm—and an uncertain world *ex ante*, with each person facing an equal chance of being able or infirm.[12] A planner might choose to arrange things so that the outcome is *ex ante* Pareto optimal. This means that it is impossible to raise one person's expected utility without lowering the other's.

Expected utilities can be written as

$$\text{EU} = \tfrac{1}{2}U^A(y_A) + \tfrac{1}{2}U^I(y_I, z). \tag{31}$$

The von-Neumann-Morgenstern functions U^A and U^I must be increasing, concave if people are risk averse, and U^I must be ordinally equivalent to U.

Since both people have identical expected utilities, an *ex ante* Pareto optimum results when EU is maximized subject to (3).[13] This procedure will choose a particular first-best *ex post* Pareto optimum, on the FBF discussed above. It is easy to construct examples where this has incentive problems *ex post*.

For example, if

$$\text{EU} = \tfrac{1}{2}\ln(y_A + e) + \tfrac{1}{2}\ln\left(y_I - e^{1-z} + e\right), \tag{41}$$

first-best *ex ante* optimally requires, for $k = 6, (\hat{y}_A, \hat{y}_I, \hat{z}) = (2,3,1)$. The utility functions used here are ordinally equivalent to the ones of previous sections. The allocation $(2,3,1)$ with $y_I > y_A$ does not satisfy the second-best constraint. Therefore, there will be a temptation for Able to pose as infirm when the uncertainty is resolved. A planner with information problems must choose second- or third-best optima.

Can insurance companies help in this case? Clearly, the usual first-best optimal insurance contracts with money payoffs and Walrasian spot prices will not do. It seems clear, however, that insurance companies can provide insurance in kind. This can be, for example, medical care that cannot be sold. Such phenomena can be observed. Insurance companies often pay contractors directly when claims on fire insurance policies are made. Indeed, this example suggests that self-selection and moral hazard are closely related. If lying about the occurrence of a fire is not profitable because companies know whether a fire occurred or not, then setting fire to an overinsured building may be the next "best" action.

---

[12] We rule out uncertainty about the number of infirm people for simplicity. If there is uncertainty of this type, then satisfaction of the self-selection constraints will reveal the proportions of the two types.

[13] This procedure treats the two individuals "anonymously." If individuals have names that are independent of health status then other *ex ante* Pareto optima can be found by awarding bundles that depend on health status and name. This occurs because there are now two possible social states—person one is able and person two infirm, and person two able and person one infirm. These Pareto optima are *ex post* Pareto optimal as well.

In second-best optima, with payments in cash rather than kind, people may well protest that they are "overinsured," preferring to transform some of their medical or other benefits into consumption goods.

V. Conclusion

This paper is about Pareto efficiency in a world of incomplete information. Second- and third-best Pareto optima are characterized when rationing (second-best) or taxes and subsidies (third-best) are used to allocate resources. Second-best optima that are not first-best occur when the first-best *distribution* is not compatible with the self-selection constraints. Thus, when the government finds that more yams may not be transferred to Infirm because Able covets them, transfers of medical care may still be feasible and useful. That is the reason why we observe an apparent "overprovision" of medical care.

A similar phenomenon occurs with third-best. Subsidized or taxed medical care changes the distribution of well-being when incomes must be equal for self-selection reasons.

These second- and third-best Pareto optima are inconsistent with many of our intuitions. Cash is no longer better than kind, markets are not superior to rationing, and "marginal willingness-to-pay equal to marginal cost" is not a good cost-benefit rule.

However, some of the second-best optima are also first-best. Would it be possible for a "Paretian" welfare economist to exhibit a preference for second-best optima that are also first-best over the other second-best optima? Such a choice would require an additional ethical criterion, one that distinguishes among Pareto optima. But the separation of the second-best frontier into two qualitatively distinct parts is a distributional phenomenon. Second-best optima that are also first-best favor Able over Infirm; therefore, a preference for these Pareto optima is, in our model, a distributional preference.

Do the results of this paper support rationing of medical care or other government financed or provided services? In our model, there is only one kind of sick person, and the government knows the preferences that each of these people has. If there are many illnesses, and if any two people with the same illness have the same preferences, then our model can be extended to this case. If different people have different preferences or incomes when they have the same illness, then the case for pure rationing is weakened. Clearly, optimal policy choices in the real world may include mixed schemes, or, possibly, two-part tariffs.[14]

APPENDIX

LEMMA: $\underline{u}_A < \bar{u}_A$.

PROOF: At $\underline{u}_A$,

$$\underline{u}_A = \underline{y}_A = \underline{y}_I. \tag{A1}$$

where $\underline{y}_I$ is the first-best allocation of yams to Infirm corresponding to $\underline{u}_A$. This means that $\underline{u}_A < k$, and, because z is normal for Infirm, $\underline{z} > 0$. Thus

$$U(\underline{y}_A, 0) = U(\underline{y}_I, 0) < U(\underline{y}_I, \underline{z}) =: \underline{u}_I, \tag{A2}$$

and this implies

$$\underline{u}_I > U(\underline{u}_A, 0). \tag{A3}$$

From (14), $\underline{u}_A < \bar{u}_A$.

---

[14] In British Columbia, physiotherapy is financed through the medical plan with a two-part tariff—five dollars per visit for the first twelve visits in a year and market price for additional visits.

PROOF OF THEOREM 1:

(i) If $u_A > \bar{u}_A$, then along the FBF ($u_I^f = F(u_A)$),

$$\text{(A4)} \qquad u_I^f = F(u_A) < U(u_A, 0).$$

Any second-best utility $u_I = u_I^s$ for $u_A > \bar{u}_A$ satisfies

$$\text{(A5)} \qquad u_I^s < F(u_A) < U(u_A, 0).$$

But this means that (13) is not satisfied, and, therefore, such a u_I^s does not exist.

(ii) If $\underline{u}_A \le u_A \le \bar{u}_A$, both self-selection constraints are satisfied along the FBF. Consequently, each first-best optimum is a second-best optimum and the second-best frontier (SBF) coincides with the FBF.

(iii) For $u_A < \underline{u}_A$, constraint (10) is not satisfied on the FBF but (11) is. Accordingly, second-best allocations maximize

$$\text{(A6)} \qquad U(y_I, z)$$

subject to

$$\text{(A7)} \qquad y_I + z \le k - u_A$$

and

$$\text{(A8)} \qquad y_I \le u_A.$$

Since (A8) binds, second-best allocations must satisfy

$$\text{(A9)} \qquad u_I^s = U(u_A, k - 2u_A).$$

Hence,

$$\text{(A10)} \qquad \frac{du_I^s}{du_A} = U_y - 2U_z.$$

At $u_A = \underline{u}_A$, second-best and first-best coincide, with

$$\text{(A11)} \qquad \frac{U_y}{U_z} = 1$$

($\underline{z} > 0$ because of normality). Consequently, from (A10) and (A11)

$$\text{(A12)} \qquad \left.\frac{du_I^s}{du_A}\right|_{u_A = \underline{u}_A} = U_y - 2U_z = -U_z < 0.$$

Hence, there are second-best Pareto optima for $u_A < \underline{u}_A$. At all these allocations, the first-order conditions (with λ and μ the Lagrange multipliers for constraints (A7) and (A8), respectively) are

$$\text{(A13)} \qquad U_y(y_I, z) = \lambda + \mu$$

and

$$\text{(A14)} \qquad U_z(y_I, z) = \lambda,$$

as long as $u_A > 0$. (A13) is satisfied with equality because z must be strictly increasing as u_A descends from $\underline{u}_A$. Hence

$$\text{(A15)} \qquad \frac{U_z(y_I, z)}{U_y(y_I, z)} = \frac{\lambda}{\lambda + \mu} < 1.$$

(A15) must hold as $u_A \to 0$ as well if a Pareto optimum exists there. Further since $y_I^f > y_A^f = u_A$ in the first-best allocation, y_I must be reduced in the second-best. Hence, z is overprovided in the second-best.

PROOF OF THEOREM 2:

(i) Since y and z are both normal goods for Infirm, the EIWE is unique. Consider any other first-best optimum. Because U is strongly quasi concave, there is only one possible choice of q ($q = 1$) as long as y_I and z are positive. This is the case for all first-best allocations with $u_A < k$. Thus incomes must differ, and these allocations are not third-best. At $u_A = k, m_I = 0$ and $m_A = k$ so this is not third-best.

(ii) Suppose that a second-best optimum is not the EIWE and is a first-best optimum. Then by (i), it is not third-best. Suppose it is second-best but not first-best. Then by Theorem 1, $y_I = y_A = u_A < \underline{u}_A$, and since $u_A < \underline{u}_A$, $z > 0$. Because U is strongly quasi concave, q is unique, and $q > 0$. We have

$$\text{(A16)} \qquad m_A = y_A < y_I + qz = m_I$$

so that it cannot be third-best.

(iii) Let the q that maximizes $u_I = V(q, u_A)$ subject to (28) and (29) be $\tilde{q}(=Q(u_A))$ and define $\tilde{Y}(u_A) := Y(\tilde{q}, u_A)$ and $\tilde{Z}(u_A) := Z(\tilde{q}, u_A)$. Then, from (29),

$$u_A + \tilde{Y}(u_A) + \tilde{Z}(u_A) = k, \tag{A17}$$

and the maximum may be written $u_I^t = T(u_A)$. This touches the FBF at the EIWE and lies below it elsewhere. Under our assumptions, derivatives of $u_I^t, \tilde{Y}, \tilde{Z}$ with respect to u_A exist in a neighborhood of the EIWE, and

$$\begin{aligned} \left.\frac{du_I^t}{du_A}\right|_{\text{EIWE}} &= \frac{d}{du_A} U(\tilde{Y}(u_A), \tilde{Z}(u_A)) \\ &= U_y \tilde{Y}'(u_A) + U_z \tilde{Z}'(u_A) \\ &= \mu(\tilde{Y}'(u_A) + \tilde{Z}'(u_A)) \\ &= \mu(-1) \\ &= -\mu, \end{aligned} \tag{A18}$$

where

$$\mu = U_y = U_z \tag{A19}$$

at the EIWE, and $\tilde{Y}'(u_A) + \tilde{Z}'(u_A) = -1$ from (A17). Along the FBF

$$\begin{aligned} \frac{du_I^f}{du_A} &= \frac{dV(1, k - u_A)}{du_A} \\ &= -V_m(1, k - u_A) \\ &= -\mu. \end{aligned} \tag{A20}$$

Thus the two frontiers are tangent with a negative slope at the EIWE, and there are third-best optima in a neighborhood.

(iv) Suppose that a third-best optimum is not the EIWE. Then, it satisfies the second-best constraints (10) and (11). It cannot be a second-best optimum because of (ii). Therefore, it is Pareto dominated by a second-best optimum.

REFERENCES

Cohen, G. A., "Self-Ownership, World-Ownership and Equality, Part II," *Social Philosophy and Policy*, May 1986, *3*, 77–96.

Diamond, Peter A. and Mirrlees, James A., "A Model of Social Insurance with Variable Retirement," *Journal of Public Economics*, December 1978, *10*, 295–336.

Diewert, W. E., Avriel, M. and Zang, I., "Nine Kinds of Quasiconcavity and Concavity," *Journal of Economic Theory*, December 1981, *25*, 397–420.

Dye, Ronald A. and Antle, Rick, "Cost-Minimizing Welfare Programs," *Journal of Public Economics*, July 1986, *30*, 259–65.

Harris, Milton and Townsend, Robert, "Resource Allocation Under Asymmetric Information," *Econometrica*, January 1981, *49*, 33–64.

Hammond, Peter, "Markets as Constraints: Multilateral Incentive Compatibility in Continuum Economies," *Review of Economic Studies*, July 1987, *54*, 399–412.

Nichols, Albert L. and Zeckhauser, Richard J., "Targeting Transfers Through Restrictions on Recipients." *American Economic Review Proceedings*, May 1982, *72*, 372–77.

Roberts, Kevin, "The Theoretical Limits to Redistribution," *Review of Economic Studies*, July 1984, *51*, 177–95.

Schmundt, Maria, Smolensky, Eugene and Stiefel, Leanna, "When Do Recipients Value Transfers at Their Costs to Taxpayers?," in *Integrating Income Maintenance Programs*, Irene Lurie, ed., New York: Academic Press, 1975.

Economic and Political Foundations of Tax Structure

By WALTER HETTICH AND STANLEY L. WINER*

The paper derives the essential elements of tax systems as the outcome of rational behavior in a model where government maximizes expected support and where opposition to taxation depends on the loss in full income. The analysis treats the level of expenditures as endogenous and integrates the influence of administration costs with that of political and economic factors. Tax structure is shown to be a system of related components in equilibrium.

While the theoretical literature on taxation has flourished in the past fifteen years, tax theory continues to suffer from important limitations. Perhaps the most serious shortcoming is a dichotomy in assumptions on what motivates public and private decisions. While private behavior is modeled as self-interested in the way common in other areas of economics, public decision makers are assumed in much of the tax literature to choose and implement policies according to general social criteria such as efficiency and equity. The conflicting treatment of private and public choices restricts the ability of economists to understand the operation of actual tax systems and to explain why they have the characteristics and structure that we commonly observe.

We show in this paper that the essential facts of tax systems can be explained as the natural outcome of self-interested decision making if such behavior is assumed in both the private and public sectors. Actual tax systems are complicated and often elaborate. Underneath their rather baroque appearance lies a simple skeleton, however, consisting of a limited number of parts or components. The main elements in all tax systems are tax bases, rate structures, and special provisions, such as exemptions and deductions. A theoretical analysis of tax structure must show how these elements arise as a result of private and public choices and what determines their design and their importance within the system as a whole.

The emphasis in the paper is on the reasons for the emergence of the tax skeleton with a given set of political institutions. An alternative approach to positive tax theory is to assume the existence of one particular aspect of tax structure and to allow voters to choose relevant parameters through majority rule (Thomas Romer, 1975; K. W. S. Roberts, 1977; A. H. Meltzer and S. F. Richard, 1981, 1983). While focusing on one feature permits more detailed analysis of that component, it avoids the broader question of why tax structure as a whole exists and how different parts of the tax skeleton are related. A similar limitation also affects work which emphasizes particular determinants, such as administration costs and opportunities to tax (Richard Musgrave, 1969).[1]

The paper starts with the presentation of a basic model in which a government maximizing expected support sets tax rates for N

---

*Department of Economics, California State University, Fullerton, CA 92634, and School of Public Administration, Carleton University, Ottawa, Canada K1S 5B6. The authors are indebted to participants in seminars at Carleton University, the Universities of Toronto, Western Ontario, Montreal and Rochester, and the Claremont Graduate School. In addition, they would like to acknowledge stimulating discussions with George Warskett and helpful comments by Peter Coughlin and referees of this *Review*. The research was supported by a grant from the Social Sciences and Humanities Research Council of Canada.

[1] Musgrave refers to tax handles, a term which appears to include the changing opportunities to levy taxes as well as tax administration costs. Some authors, such as Harley Hinrichs (1966) also rely on exogenous cultural and political factors. For an attempt to base the explanation of tax structure primarily on political factors, see Susan Hansen (1983).

individuals who have different economic and political responses but who engage in the same type of economic activity. The resulting equilibrium serves as a reference solution for later parts of the paper where the analysis is generalized to include many activities by taxpayers and to account for the creation of bases, rate structures, and special provisions. The paper also contains a discussion of the relationship of tax rates and tax revenues and a comparison with two well-known alternative approaches, optimal taxation and the Leviathan model.

I. A Basic Model

The approach to political economy adopted here relies on the modeling of political equilibrium rather than of the political process. There is a successful literature that uses models of this type, including work by George Stigler (1971), Sam Peltzman (1976), and Gary Becker (1983), and having its roots in the writings of Joseph Schumpeter (1950) and Anthony Downs (1957). This literature has focused on the implications of different equilibrium outcomes for government policy while bypassing questions concerning the existence and stability of political equilibrium and the explicit derivation of government objective functions in a game-theoretic context. Political equilibria are interpreted as the outcome of a competitive process and no independent role is assigned to bureaucracy.[2] Our paper can be viewed as a further formalization and extension of this literature.

A second important tradition of analysis has taken a quite different direction, emphasizing theoretical questions of existence and stability (see for example, Peter Ordeshook, 1986, ch. 4). This tradition has not been concerned with developing institutional implications, perhaps because problems of existence and stability are sufficiently difficult in their own right. While we see no general solution emerging as yet from this literature which could serve as a basis for modeling tax structure, we will suggest briefly how our objective function could be derived from the solution to a particular electoral game described in a recent paper by Peter Coughlin (1986).

We assume that the government's objective in designing a tax structure is to maximize expected support, a term which can be interpreted in both a narrow and broad manner. In the "narrow" version, individual support for the government depends (*i*) on the benefits from public goods and the loss in full income resulting from taxation and (*ii*) on characteristics which determine how a particular individual's net economic benefit from the fiscal system is translated into a probability of voting for the government. Examples of relevant such characteristics include the cost of voting, age, and the taste for civic duty. The government's objective in this model can be interpreted as the maximization of expected votes, an objective often used in the literature since Downs. Maximization of expected votes can be regarded as a simple way of capturing the motivation of a government which is unsure of the identity or characteristics of its opponents in future elections (Arthur Denzau and Michael Munger, 1986; David Mayhew, 1974).

A broader interpretation of the term support which we have found useful in empirical research (Walter Hettich and Stanley Winer, 1984; Winer and Hettich, 1987) is also possible. In this view, support has a third component in addition to those already mentioned. As seen by the government, effective support depends not only on the likelihood that an individual will cast a favorable vote in the next election but also on the individual's relative political influence in a world where equality of franchise differs from equality of influence. As a result, the government maximizes the weighted sum of expected votes where weights depend on voter characteristics such as interest group membership and strength and on individual attributes such as personal wealth.

In the model presented below, the probability of any individual's voting for the government is influenced positively by the services received from a pure public good G and negatively by his loss in full income v (including deadweight loss) from taxation. Voters base their decision on whether to support the government on how they are affected by benefits and taxes and are not influenced by how others are treated. Moreover, individual taxpayers see no connection between the level of services provided and their own tax burden. This implies that there is no direct link between expenditure structure and tax structure, even though the level of expenditures is endogenous and affects tax structure indirectly through the government's budget constraint. The separation of taxes and expenditures is an important characteristic of modern fiscal systems, and for this reason, a good starting point in constructing a positive theory of tax structure. One may also note that assuming such separation helps to simplify a rather complex theoretical problem and has proven useful in empirical research.

---

[2] The underlying assumptions for this approach are that political competition is sufficiently strong to force convergence to an optimal strategy and that the ensuing equilibrium is stable.

Given the above assumptions, we can represent the ith voter's support for the government as:

$$(1) \quad \{b_i(G)-c_i(v_i)\} \quad i=1,2,\dots,N$$

where N is the number of taxpayers and where $\partial b_i/\partial G>0$ and $\partial c_i/\partial v_i>0$. In the narrow view, (1) normalized appropriately represents the probability of voting for the government, while in the broader interpretation it is that probability weighted by additional factors determining effective political influence.[3] The term b_i in (1) represents benefits for individual i while c_i can be interpreted as effective opposition to taxation by the rational voter. The greater is c_i, the smaller is the expected support from the ith voter. Opposition depends positively on the loss in full income v_i which may be written as

$$(1a) \qquad v_i=T_i+d_i,$$

where T_i is the tax payment and d_i is the deadweight loss or welfare cost of taxation for the ith voter.

We shall also assume that taxation of the ith voter is proportional at rate t_i, that B_i, the level of the taxable activity of the ith voter, is related negatively to t_i, while the welfare loss d_i depends positively on t_i and on exogenous factors x_i that determine the nature of the supply function for the taxable activity. (If the taxable activity is work, for example, an important component of x_i will be the taste for leisure.) Hence for $i=1,2,\dots,N$,

$$(1b) \qquad T_i=t_i\cdot B_i,$$

$$(1c) \qquad B_i=B_i(t_i,x_i)\ \partial B_i/\partial t_i<0$$

and

$$(1d) \qquad d_i=(t_i,x_i)\ \partial d_i/\partial t_i>0$$

with $d_i=0$ and $v_i=T_i$ if taxes did not disturb behavior.[4] Equations (1c) and (1d) reflect the voter's utility-maximizing response to taxation. For simplicity it is assumed that all taxpayers engage in the same type of activity (but that they are otherwise unique). and that their economic and political responses to taxation are known to the government without cost.

The government chooses the level of public expenditure G and tax rates $t_1,t_2,\dots,t_n$ so as to maximize expected support

$$(2) \qquad \sum_{i=1}^{N}\{b_i(G)-c_i(v_i)\},$$

---

[3] One approach to the narrow interpretation of (1) is to define an index of support $I_i=b_i(G)-c_i(v_i)$. Then let p_i be an increasing function of I_i which translates the support index into a voting probability, $0\le p_i(I_i)\le 1$. The standard logit or probit model could be used to define p_i.

[4] Putting deadweight costs (1d) explicitly into the model is a convenient method of incorporating the full consequences of taxation for individual welfare. See Becker (1983) for a similar construction.

subject to the government budget restraint

$$G - \sum_{i=1}^{N} t_i \cdot B_i = 0, \tag{2a}$$

and subject to taxpayers' responses to taxation reflected by equations (1c) and (1d).

While we shall not further develop this aspect of the analysis, it is of interest to note that maximization of (2) can be viewed as describing the outcome of a particular two-candidate game. Coughlin (1986) has constructed such a game where every voter has a finite probability of voting for any one of the two candidates. Equilibrium can be represented by the maximization of any candidate's expected vote over the policy space available to him.[5] Candidates use a binomial logit model to determine the probabilities that they will receive the support of particular voters. These probabilities and hence the candidate's expected vote are related positively to the net effect of government policy on the welfare of voters. Since we can write (2) more generally as

$$\sum_{i=1}^{N} p_i \{ f_i(G, \underline{t}) \}, \tag{2'}$$

where p_i is the ith voter's probability of voting for the government and f_i represents the effect of fiscal policy on his welfare (f depends positively on public services G and negatively on tax rates $\underline{t}$), maximization of (2) or (2′) and the equilibrium of the voting game characterize the same policy choices.[6]

In the basic model, the first-order conditions for a solution to the government's problem consist of equations (2a), (3a), and (3b):

$$\sum_i \partial b_i / \partial G - \lambda = 0, \tag{3a}$$

$$-(\partial c_i / \partial v_i \cdot \partial v_i / \partial t_i) + \lambda(B_i + t_i \cdot \partial B_i / \partial t_i) = 0 \qquad i = 1, 2, \ldots, N. \tag{3b}$$

The first term in brackets in (3b) is the growth in opposition that results from increasing t_i. The second term in brackets represents $\partial T_i / \partial t_i$, the additional revenue raised by this rate increase. The meaning of these results can be understood more easily if we restate the first-order conditions (3) for each i as

$$\frac{\partial c_i / \partial v_i \cdot \partial v_i / \partial t_i}{B_i \cdot (1 + \varepsilon_i)} = \lambda, \tag{4}$$

where $\varepsilon_i = \partial B_i / \partial t_i \cdot t_i / B_i$ is the elasticity of the ith taxpayer's activity with respect to his tax rate.[7] Thus the politically optimal tax structure in the basic model requires a choice of tax rates that equalizes marginal political costs per dollar of additional revenue across all taxpayers. This tax structure will finance a total expenditure such that the marginal political benefit of another dollar of expenditure λ is equal to the common marginal political cost per dollar of additional revenue.[8]

---

[5] Coughlin states his results in terms of the maximization of expected plurality. However, his assumptions about candidates' beliefs concerning voter behavior imply that the maximization of expected plurality and of expected votes are equivalent objectives. One should note that in Coughlin's game, candidates propose after-tax incomes subject to a constraint on the sum of their incomes. This can also be viewed as the proposing of tax payments subject to a government budget constraint.

[6] We can think of f as the voter's post-fisc utility relative to his pre-fisc utility, the level of which is assumed given. (2′) could also be written as $\Sigma p_i \{ U_i(G, \underline{t}) - \hat{U}_i \}$, where $\hat{U}_i$ is pre-fisc welfare.

[7] We only consider interior solutions where everybody's support is solicited to some extent. This is consistent with the type of equilibrium considered in the probabilistic voting literature referred to above, where in equilibrium every voter has a positive probability of voting for each candidate.

[8] Second-order conditions for a unique solution to the government's problem include, in addition to assumptions about derivatives made above, (*i*) $b_i^{GG} < 0$, (*ii*) $v_i^{t_i t_i} > 0$, (*iii*) $c_i^{v_i v_i} > 0$, and (*iv*) $B_i^{t_i t_i} < 0$. Superscripts denote partial derivatives. Note that the separation of expenditures and taxes and the independence of taxpayers assumed in the text implies that cross-partial derivatives of the objective function with respect to

While the first-order conditions above integrate economic and political behavior, they yield only a very simple tax structure which still misses several essential elements of observed tax systems. Tax structure in (4) consists of N rates on one activity, with each taxpayer being taxed at a unique rate. As yet, voters are not grouped into rate brackets, activities are not grouped into bases, and there are no special provisions. In subsequent sections we extend the basic model to account for these additional elements.

We are interested primarily in establishing a set of sufficient conditions for the existence of a stylized tax structure. Since we can accomplish this without endogenizing interest group formation and without relaxing the assumption of independence among net political benefit functions of voters, or the assumption that their economic and political responses to taxation are known to the government without cost, we shall maintain all three assumptions throughout the paper.

II. Taxation of Many Activities

In this section we demonstrate that the taxation of many activities is a natural outcome of expected support maximization. The loss in full income given by (1a) can be generalized to $v_i = v_i(T_{i1}, T_{i2}, \ldots, T_{iJ}, d_i)$ in the case in which each taxpayer conducts J activities and faces tax rates $t_{i1}, t_{i2}, \ldots, t_{iJ}$. Equations (1c) and (1d) may also be similarly generalized. In this case, the government budget restraint (2a) becomes

$$(2a') \qquad G - \sum_{i=1}^{N} \sum_{j=1}^{J} t_{ij} \cdot B_{ij} = 0,$$

and the first-order conditions for the government's problem (4s) change to

$$(4') \qquad \frac{\partial c_i/\partial v_i \cdot \partial v_i/\partial t_{ij}}{B_{ij} \cdot (1+\varepsilon_{ij}) \sum_{h \neq j} t_{ih} \cdot \partial B_{ih}/\partial t_{ij}} = \lambda$$

$$i = 1, 2, \ldots, N; \; j, h = 1, 2, \ldots, J$$

where $\varepsilon_{ij} = \partial B_{ij}/\partial t_{ij} \cdot t_{ij}/B_{ij}$. The second term in the denominator on the left side of (4′) represents the effect of taxing activity j on other activities conducted by a given taxpayer.[9] Activities of different taxpayers are, however, assumed to be independent in this formulation.[10]

Equation (4′) generalizes the conditions stated in (4). It implies that the politically optimal tax structure requires marginal political opposition per dollar of tax revenue to be equalized across taxable activities for each taxpayer, as well as to be equalized across taxpayers for each activity.

The above argument indicates that the evolution of tax structure is closely related to economic change and development. Minimization of opposition to taxation requires the adjustment of tax structure whenever the broad nature of activities conducted by taxpayers changes. The same argument also explains why tax structure is complex—disregarding differences in the welfare consequences of taxation across activities conducted by any taxpayer or across all taxpayers increases opposition to taxation because such disregard makes voters worse off. This may be the reason why tax simplification remains elusive although it appears to be universally endorsed as a good idea.

Equation (4′) indicates that complexity in tax structure is politically rational. In fact, it implies even greater complexity than is usually observed. The task is thus not only to justify the existence of complexity, but also to delineate its limits.

---

policy instruments are zero. It is also assumed that taxable activities are independent across taxpayers. Conditions (*ii*) and (*iii*) imply that marginal opposition to taxation increases with tax rates. Conditions (*i*) to (*iii*) ensure the objective function (2) is strictly concave, while (*iv*) ensures that the constraint (2a) is strictly convex.

[9] If taxpayer i engages in J activities, equation (1c) must be reformulated as $B_{ij} = B_{ij}(t_{i1}, t_{i2}, \ldots, t_{iJ}, x_i)$.

[10] This last assumption is consistent with the simplification introduced earlier that net political benefit functions (1) are independent across taxpayers.

III. Grouping and the Choice of Rate Structures and Bases

In the solution to the government's problem in Section II we have N taxpayers, J activities, and $N \times J$ tax rates. This is unrealistic in two respects. First, activities are generally grouped into bases which consist of similar or related activities. In addition, taxpayers are sorted or grouped into rate brackets where despite interpersonal differences they pay the same tax rates.

Deviations from unique treatment of activities or of taxpayers will cause a loss in political support. What must be explained is why the government decides to accept such losses, that is, what offsetting advantage can be gained in exchange for the grouping. The answer lies in reduced administration costs. Resources released in this way can then be used to provide more public goods and therefore to obtain increased support. The government's problem is to balance the marginal loss in support from grouping with the marginal gain in support from spending resources not used in administration.

Since we developed the basic model using a framework with N taxpayers and one activity, it is convenient to begin the analysis by considering the sorting of taxpayers into rate brackets on one activity. We shall indicate later how the solution can also be interpreted as the rationale for combining different activities into bases.

Creation of rate brackets will mean that groups of individuals with differing levels of economic activity will be subject to the same tax rate. The government's problem is (i) to establish the politically optimal number of brackets and (ii) to assign individuals to these brackets in a manner that is consistent with its political objective.

The second part of the problem, that is, the assignment problem, is considered at length in the Appendix. We let the number of rate brackets or groups be fixed at some number $K < N$, the number of individuals. Levying the same rate on all members of a given group rather than taxing them at their unique politically optimal rates defined by first-order conditions analogous to (4) must result in a loss in expected support. If we linearize the N first-order conditions in the absence of grouping, we can show that the loss from grouping N unique individuals into K-rate brackets is minimized when taxpayers are assigned to brackets so as to minimize the within group variation in politically optimal rates.

By considering solutions of the sorting problem for different values of K, we can construct the "marginal tax discrimination" curve AA in Figure 1. For each K, this curve shows the maximum reduction in opposition possible from increasing the number of groups by one while simultaneously resorting individuals among the $K+1$ groups in the manner described above. AA lies above the horizontal axis when $K < N$ and intersects the horizontal axis at $K = N$, where all taxpayers are treated uniquely. It is assumed to decline continuously as K approaches N.

The optimal number of rate brackets K^* is shown in Figure 1 where AA intersects BB, the "marginal tax administration curve," reflecting the opportunity cost to the government of treating individuals differently when this creates administration costs.[11] For each K, BB shows the fall in political support resulting from an increase in the number of rate brackets from K to $K+1$ and a reduction in public services by an amount equal to the corresponding growth in administration costs. When λ in (4) is constant, BB will slope upward as long as the increase in administration costs rises with K when another tax instrument is added.[12]

The general solution when each taxpayer is taxed on all of his J activities involves four elements: the reduction in votes lost

---

[11] Administration costs include: (i) the cost of processing tax payments, (ii) the cost of monitoring compliance and enforcing tax codes, (iii) the cost of coordinating administrative personnel, and (iv) the costs of acquiring knowledge about taxpayers' characteristics. The last type of cost is not generally incorporated into the analysis in this paper except that lump-sum taxation is considered infeasible.

[12] It is possible that administration costs depend on the nature of the tax instruments employed, as well as on their number, as in W. P. Heller and Karl Shell (1974). We do not explore that possibility in this paper. We also rule out possible discontinuities and nonconvexities introduced by administration costs.

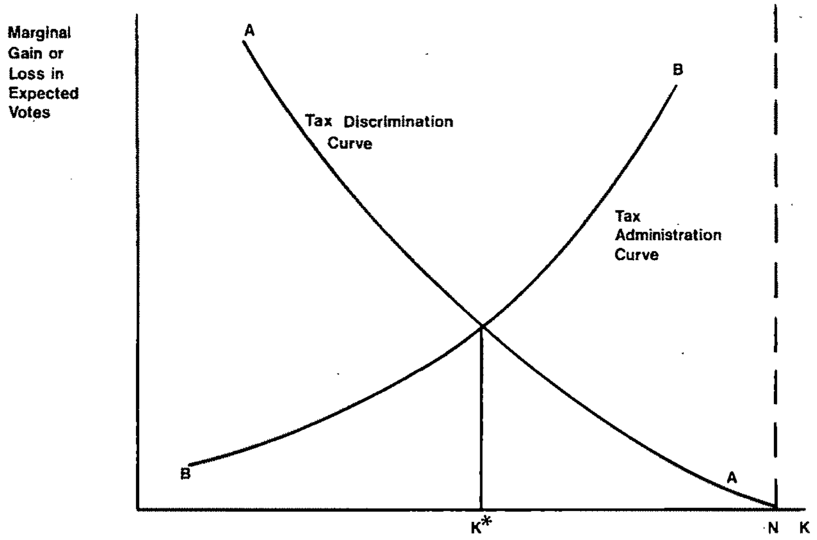


FIGURE 1. THE POLITICALLY OPTIMAL NUMBER OF RATE BRACKETS

(i.e., the marginal gain in support) when the number of rate brackets k_j on any activity is increased $\partial S/\partial k_j$, the resulting increase in administration costs $\partial A/\partial k_j$ and in total revenue $\partial R/\partial k_j$, and the additional support from spending one more dollar on public goods λ. In equilibrium,

$$\frac{\partial S/\partial k_j}{\left(\partial A/\partial k_j - \partial R/\partial k_j\right)} = -\lambda \qquad j = 1, 2, \ldots, J \tag{5}$$

that is, the marginal reduction in opposition per "net" dollar of administration costs must be equal across activities.[13]

As pointed out, the analysis has a second important application to the grouping of activities into tax bases. Consider a model with one representative taxpayer who engages in J different activities. The government can save on administration costs by combining related activities into a limited number of bases, such as occurs when incomes from different labor activities are included in the same tax base. Grouping leads to an increase in political opposition in this case since it will raise a taxpayer's deadweight loss associated with any given tax payment. On the other hand, the government receives additional support by spending resources saved in administration on the provision of public goods. The solution illustrated in Figure 1 again describes the nature of the equilibrium where K^* now refers to the number of bases for each individual rather than to rate brackets.

The preceding analysis for the first time formally integrates administration costs into political optimization.[14] In doing so, it provides a new basis for understanding the evolution of tax systems. In his pioneering work on fiscal systems, Musgrave (1969) argues that tax handles are crucial for explaining the formation and growth of tax structure. If we conceive of the term "tax handles" in a broad sense that includes opportunities to tax and to escape taxation as well as administration costs, the theory presented here formalizes what so far has mainly been implicit. In an analysis over time, changes in economic activities and in the conditions under which they are carried out will be of crucial importance together with administration costs since the nature of such activities and conditions determines the characteristics according to which the government sorts taxpayers to create bases, rate structures, and (as argued below) special provisions. The development of tax structure results therefore from an interaction between the changing ways in which people work, transact, and consume on the one hand, and the cost of administering the collection of revenues on the other. To this is added a third element, namely the influence of those factors that determine how benefits from public goods and losses in full income from taxation are translated into political action.

---

[13] First-order conditions (5) can be derived by choosing G and $(k_1, k_2, \ldots, k_J)$ to minimize the loss in support from grouping N taxpayers into k_j rate brackets for each of J activities, that is, $\min \Sigma_i \{b_i(G^*) - b_i(G)\} + S(k_1, k_2, \ldots, k_j; \underline{t}^*)$, subject to $R(k_1, \ldots, k_J) = G + A(k_1, \ldots, k_J)$, where S is decreasing in each k_j and $(G^*, \underline{t}^*)$ represent the optimal G and the vector of $N \times J$ tax rates defined in the absence of administration costs. The first term in the objective function is the loss in support from the revenue implications of grouping relative to the $(G^* \underline{t}^*)$ solution. $S(\cdot)$ represents the minimum increase in opposition to taxation from (optimally) grouping taxpayers into rate brackets, as discussed in the Appendix.

IV. Special Provisions

The final feature of the skeleton that remains to be explained is the existence of special provisions such as exemptions, deductions, and tax credits. Consider the grouping of activities into tax bases in the case with N taxpayers, each of whom engages in J activities. According to the preceding analysis, the government will create separate bases for each of the N taxpayers. The composition of bases will differ among taxpayers unless there is an additional cost, not considered so far, in administering different bases for each individual. Such costs introduce a further constraint on the grouping process leading to bases that coincide for large numbers of individuals and in rate structures that are defined on bases rather than on separate activities.

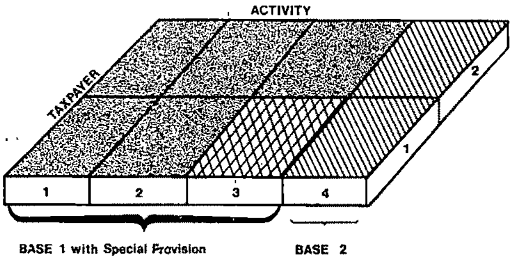


FIGURE 2. SPECIAL PROVISIONS IN TAX STRUCTURE

The argument is illustrated in Figure 2 which is drawn for two taxpayers, each of whom engages in four activities. When there are no additional costs in having separate bases for each taxpayer, the grouping process results in person 1's being taxed on two bases consisting of activities 1 and 2 and activities 3 and 4, respectively. For person 2, the two bases will consist of a combination of the first three activities and of activity 4. When there are costs to having separate bases for each individual, it may be preferable to define the first base for both individuals to include 1, 2, and 3 while making activity 4 into a second base.[15] However, it will be politically undesirable to tax activity 3 in the same way for both taxpayers. A special provision such as an exemption or a deduction in the first base will allow the government to differentiate tax treatment of activity 3 depending on the taxpayer involved.

The argument can also be approached from a second angle. The general solution in (5) specifies a different rate structure for each activity. This, no doubt, would be administratively costly. It may be preferable to define rate structures across bases comprising several activities, but to introduce some differentiation in the tax treatment of each activity in any base by having special provisions that are specific to particular activities.[16]

---

[14] See Shlomo Yitzhaki (1979) for a consideration of administration costs in an optimal tax setting. Yitzhaki does not explore the implications of heterogeneity.

[15] An additional base generally requires a new set of collection points and separate administrative arrangements.

We have now shown that the government's optimizing behavior generates all essential elements of tax structure. The skeleton is complete. The analysis also demonstrates that all the parts are interdependent. This is an important point since tax policy or tax reform often focuses only on one aspect of the system without taking account of the repercussions that must follow intervention in other parts. Rate structures, bases, and special provisions are all determined jointly. The government will furthermore try to establish a new political equilibrium each time there are shocks to the system such as changes in the factors (x_i) determining the supply of taxable activities, or in the nature of heterogeneity among taxpayers. The analysis strongly suggests that tax systems should be studied as integrated systems of essential elements and not merely as collections of unrelated or ill-designed components.

V. The Relationship Between Tax Rates and Tax Revenues

The model developed in this paper throws new light on the much discussed relationship between tax rates and tax revenues. In the basic model of Section II, this relationship will differ for each individual. While each long-run "Laffer" curve may have a backward-bending portion, political optimization in the basic model precludes tax rates which push taxpayers onto that portion, provided that political opposition increases continuously with tax rates and that the vote productivity of additional public expenditure is positive.

Opposition to taxation will always increase with tax rates if both terms in the numerator of (4) are positive. It seems reasonable to assume this to be the case, even though increases in tax rates may lower tax revenues after some point (Geoffrey Brennan et al., 1984). Then, as long as $B_i > 0$ and $\partial b_i / \partial G > 0$, the first-order conditions (4) require that ε_i exceed -1. In other words, choice of a tax rate placing a voter on the backward-bending portion of his Laffer curve would imply that the government is foregoing revenues which could be used to generate further support and that the affected voter is at the same time opposing the government more strongly than he would at lower rates.[17]

In a world of heterogeneous taxpayers and positive administration costs, this conclusion may no longer hold for some individual taxpayers. When grouping occurs, the original optimal conditions no longer apply, with the result that some individuals may become subject to a group rate placing them on the backward-bending portion of their individual rate-revenue curve. This will occur if assignment to any other group would lower the government's overall expected support. (Two factors are relevant in assigning individuals to groups: revenue collected, which can be turned into political support through production of public goods, and the expected support from the individual himself. Assuming the number of groups to be fixed, the government will assign individuals so as to maximize net support from these two sources.)

---

[16] There may be further reasons for special provisions if there is a more direct link between the expenditure and tax sides. Expression (1) assumes that expected support by a particular individual is separable in G and v_i. In actual tax systems, we do observe some provisions which may represent a more direct link between the two sides of the budget. One can imagine a government creating a tax structure based initially on (1) and then amending it subsequently for certain broad classes of individuals having similar evaluations of public output. For example, people over 65 years of age may be given a special exemption to acknowledge a lower evaluation of public services. We would expect the cost of adjusting the definition of tax bases to exceed the cost of creating special provisions. As a result, the link between expenditures and taxes—as far as it exists—will be expressed primarily through the introduction of additional special provisions.

[17] It should be pointed out that there is no difference in the time horizon of voters and the government in the model. For a discussion of the Laffer curve that focuses on such differences, see Buchanan and Dwight Lee (1983).

The argument shows that we may expect to observe some individuals whose tax payments would increase if *they alone* faced lower tax rates. One should note, however, that the same conclusion does not apply to groups. No group will be on the backward-bending part of its aggregate rate-revenue relationship. Otherwise, the government could collect more revenues from the group as a whole by lowering the group's rate, while at the same time reducing opposition from all of its members.

VI. Comparison with Other Approaches

It is useful to compare the implications of our analysis to results derived from models that are based on different premises. We shall comment on the two approaches that have been most influential in recent years, the Leviathan model and the theory of optimal taxation.

In their work on taxation, Brennan and James Buchanan (1980) put major emphasis on underlying political forces while also making the level of expenditures endogenous. Nevertheless, their approach differs in an essential aspect from ours; they are concerned with a government attempting to maximize revenues and facing no significant political constraint. The government in our model has a different objective function and is effectively constrained by political forces even to the extent that opposition by unorganized voters will be reflected in tax structure.[18]

Brennan and Buchanan do not address the problem of how tax structure is created when bases as well as other major structural elements are endogenous. Presumably, Leviathan would also combine activities and individuals into groups, but the sorting would differ in two essential respects. Grouping in our framework reflects political factors determining how losses in full income are translated into political action in addition to the abilities of taxpayers to escape taxation which are Leviathan's major concern. Furthermore, the government has an incentive to economize on administration costs, while no such motivation exists in a world where government is not constrained politically in setting tax policy.

Both these differences have implications for empirical work. Statistical research based on our model requires a more explicit treatment of the $c(v)$ function and thus the development of proxy variables for the determinants of political influence. In addition, it is necessary to formulate variables reflecting administration costs (Winer and Hettich, 1987). Presumably, no such variables would be included in a model based on the existence of Leviathan-type behavior. Wallace Oates (1985) investigates such behavior by looking at whether or not decentralization constrains the size of government. Our theory suggests a more direct test, namely whether the evolution of tax structure and revenue composition is influenced by political and administrative factors in addition to economic ones.

While there are formal similarities between the theory of optimal taxation (OT) and the analysis developed here, there are also important substantive differences. In OT, the government maximizes a welfare function, usually written as a weighted sum of individual utilities, with the weights reflecting exogenously given distributional preferences. To our knowledge, writers on OT have not dealt with the creation of tax structure when tax bases are endogenous. While their work could be extended to include an analysis of sorting, the resulting equilibrium would differ since factors determining how losses in full income are translated into political support have no role in OT.

Optimal taxation is generally conceived of as a normative theory not concerned with explaining actual government behavior, but intended to provide a standard of reference which abstracts from the political setting in which tax policy is made. In this context, the interesting comparison concerns the efficiency of equilibrium.

---

[18] For this reason, our model also differs from that of Becker (1983) in which only the interests of organized groups are reflected in policy choices by the government.

If the optimal tax problem is stated as the maximization of social welfare subject to the raising of a fixed budget, efficiency in taxation requires that the change in social welfare per dollar of additional revenue be equalized across revenue sources. Efficiency of this kind is not a general characteristic of the model presented here, where as in equation (4), a support-maximizing government equalizes the change in expected political support per dollar of additional revenue across revenue sources. As a result, support-maximizing governments will create tax structures that differ significantly from solutions envisioned in the OT literature.

We may still ask whether a tax system of the kind described in the paper can be considered globally efficient. If we interpret tax structure as the long-run equilibrium of a competitive political system in which political opposition depends on the loss in full income, no political party can offer an alternative tax system generating the same political support with a lower welfare loss for any individual. In this sense tax structure is efficient for the existing set of political institutions. This does not mean that an alternative set of institutions could not yield a better tax system. The argument does, however, direct debate on tax reform toward the redesign of political institutions.

VII. Concluding Remarks

Existing tax systems are composed of a limited number of basic elements which have been combined to form complicated structures. To understand why tax systems have the appearance and characteristics that we observe, we must explain why the basic elements are used as building blocks and why they are combined in particular ways.

The paper demonstrates that the essential stylized facts of observed tax systems can be viewed as the outcome of optimizing political and economic behavior. It further shows that the way in which these elements are combined into different structures depends on administration costs and on the nature of political and economic responses to taxation among individuals.

Tax structure is a system of related parts in equilibrium, not merely a collection of separate and ill-designed components. This has important implications for the understanding of tax policy. Changes must pass a political as well as an economic test and reforms in one part of the system may lead to unexpected repercussions elsewhere as the government attempts to establish a new equilibrium. The analysis also suggests that the evolution of tax systems can be viewed as a sequence of responses to changing economic, administrative, and political factors. Future empirical work should adopt a framework that can account for the systematic influence of these determinants on tax history.

APPENDIX: GROUPING OF TAXPAYERS INTO RATE BRACKETS (OR ACTIVITIES INTO BASES)

Let the number of rate brackets be some number $K < N$, the number of taxpayers, and assume that first-order conditions in the absence of grouping can be written as

$$a_{ik} + m \cdot t_{ik} = g_{ik} - h \cdot t_{ik}. \tag{A1}$$

The left side of (A1) represents the marginal political cost of taxing person $i(i = 1, 2, \ldots, n_k)$ in group $k(k = 1, 2, \ldots, K)$, analogous to the numerator in (4), while the right side stands for the marginal benefit of raising t_{ik} and spending the extra revenue. Heterogeneity of behavior is captured by differences in the constant terms. Imposing a rate t_{ik}, differing from the politically optimal rate t_{ik}^* found by solving (A1), results in a loss in support from each taxpayer equal to the integral of the difference between marginal political costs and benefits over the interval from t_{ik}^* to t_{ik}. Given (A1), this integral is equal to $v \cdot (t_{ik} - t_{ik}^*)^2$ where $v = (m + h)/2$. Similarly, the loss in support from taxing all members of any group k alike at rate t is equal to $v \cdot \Sigma_i (t - t_{ik}^*)^2$. The rate t that minimizes this loss is the least squares solution $t_{\cdot k} = \Sigma_i t_{ik} / n_k$. Thus with K brackets or groups the total loss in support from grouping can be reduced to $v \cdot \Sigma_k \Sigma_i (t_{\cdot k} - t_{ik}^*)^2$. This loss can be made as small as possible if taxpayers are assigned among groups so as to minimize the variation of the t_{ik}^*'s within each group.

REFERENCES

Atkinson, A.B. and Stiglitz, J.E., *Lectures on Public Economics*, New York: McGraw-Hill, 1980.

Becker, Gary S., "A Theory of Competition Among Pressure Groups for Political Influence," *Quarterly Journal of Economics*, August 1983, *98*, 371–400.

Brennan, Geoffrey and Buchanan, James M., *The Power to Tax*, Cambridge MA: Cambridge University Press, 1980.

______, **Bohanon, Cecil and Carter, Richard**, "Public Finance and Public Prices; Towards a Reconstruction of Tax Theory,"*Public Finance*, 1984, *39*, 157–79.

Buchanan, James and Lee, Dwight, "Politics, Time and Laffer Curve," *Journal of Political Economy*, August 1982, *90*, 816–19.

Coughlin, Peter J., "Elections and Income Redistribution," *Public Choice*, 1986, *50*, 27–91.

Denzau, Arthur T. and Munger, Michael C., "Legislators and Interest Groups: How Unorganized Interests Get Represented," *American Political Science Review*, March 1986, *80*, 89–106.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper & Row, 1957.

Hansen, Susan B., *The Politics of Taxation: Revenue Without Representation*, New York: Praeger, 1983.

Heller, W.P. and Shell, Karl, "On Optimal Taxation with Costly Administration," *American Economic Review Proceedings*, May 1974, *64*, 338–45.

Hettich, Walter and Winer, Stanley L., "A Positive Model of Tax Structure," *Journal of Public Economics*, June 1984, *24*, 67–87.

______ **and** ______, "Blueprints and Pathways: The Shifting Foundations of Tax Reform," *National Tax Journal*, December 1985, *38*, 423–45.

______ **and** ______, "Federalism, Special Interests and the Exchange of Policies for Political Resources," *European Journal of Political Economy*, 1987, *3*, Special Issue, 33–54.

Hinrichs, Harley H., *A General Theory of Tax Structure Change During Economic Development*, *Cambridge*, *MA*: *Harvard Law School*, 1966.

Mayhew, David R., *Congress*: *The Electoral Connection*, New Haven: Yale University Press, 1974.

Meltzer, A.H. and Richard S.F., "A Rational Theory of the Size of Government," *Journal of Political Economy*, October 1981, *89*, 914–27.

______ **and** ______, "Tests of a Rational Theory of the size of Government," *Public Choice*, 1983, *41*, 403–18.

Musgrave, Richard A., *Fiscal Systems*, New Haven: Yale University Press, 1969.

Oates, Wallace E., "Searching for Leviathan: An Empirical Study," *American Economic Review*, September 1985, *75*, 748–57.

Ordeshook, Peter C., *Game Theory and Political Theory*: *An Introduction*, Cambridge, MA: Cambridge University Press, 1986.

Peltzman, Sam, "Toward a More General Theory of Regulation," *Journal of Law and Economics*, August 1976, *19*, 219–40.

Roberts, K.W.S., "Voting over Income Tax Schedules," *Journal of Public Economics*, December 1977, *8*, 329–40.

Romer, Thomas, "Individual Welfare, Majority Voting, and the Properties of a Linear Income Tax," *Journal of Public Economics*, February 1975, *4*, 163–85.

Schumpeter, Joseph A., *Capitalism*, *Socialism*, *and Democracy*, 3rd ed., New York: Harper & Row, 1950.

Stigler, George J., "The Economic Theory of Regulation," *Bell Journal of Economics*, Spring 1971, *3*, 3–21.

Winer, Stanley L. and Hettich, Walter, "The Evolution of Revenue Systems: A Theoretical and Empirical Investigation," Carleton Economics Papers, No. 87-04, May 1987.

Yitzhaki, Shlomo, "A Note on Optimal Taxation and Administration Costs," *American Economic Review*, June 1979, *69*, 475–80.

Plea Bargaining and Prosecutorial Discretion

By JENNIFER F. REINGANUM*

A model of plea bargaining with asymmetric information is presented. The prosecutor's private information is the strength of the case; the defendant's is his guilt or innocence. In equilibrium, some cases are dismissed because they are too likely to involve an innocent defendant. In the remaining cases, the prosecutor's sentence offer reveals the strength of the case. A particular restriction on prosecutorial discretion is shown to be welfare-enhancing for some parameter configurations.

Under current practice, prosecutors have essentially unlimited discretion to dismiss a case or to negotiate a guilty plea to a lesser crime, thereby guaranteeing a lighter sentence.† Much controversy surrounds the exercise of prosecutorial discretion. While it is acknowledged that guilty pleas save resources which would otherwise be devoted to trials, one major concern of opponents to plea bargaining is that the prosecutor is in an unfairly strong bargaining position. One source of bargaining strength for the prosecutor is simply the fact that the defendant is required to deal with him, and cannot "shop" for a better deal.

Another asymmetry between the bargaining positions of the prosecutor and the defendant stems from the fact that the prosecutor typically has better information about the strength of the case than does the defendant. The prosecution has presumably interviewed witnesses (including the defendant) and gathered evidence. In the modal criminal case, the time and investigative resources available to the defense are a fraction of those available to the prosecution. Thus although in principle the defense may have equal access to evidence and witnesses,[1] in practice it relies heavily on summary data from the prosecution. It seems likely that in this case there will remain some uncertainty on the part of the defense regarding the strength of the prosecution's case. Similarly, despite the evidence which the prosecution may possess, it seems likely that there will remain some uncertainty on the part of the prosecution regarding the defendant's factual guilt.

In this paper, a model of the plea-bargaining process in the presence of asymmetric information is developed.[2] While the issue of

---

*Departments of Economics and Management Science, University of Iowa, Iowa City, IA 52242. I would like to thank Kim Border, Richard Craswell, Ted Groves, Barry Nalebuff, Herman Quirmbach, Eric Rasmusen, and two anonymous referees for helpful comments and discussion. The financial support of the National Science Foundation, grant no. SES-8710578, the Alfred P. Sloan Foundation, and the Graduate School of Business of the University of Chicago are gratefully acknowledged.

†Plea bargaining is a prominent feature of the criminal justice system. According to Alschuler (1981, p. 652), "it is commonly estimated that 90% of all criminal convictions are the result of guilty pleas." Moreover, a large fraction of the cases are simply dismissed; based upon a sample of 1,382 felony arrests in New York City in 1971, the Vera Institute of Justice (1977, p. 7) estimates that "43% of the felony arrests were disposed of by dismissal."

[1] In a related paper, Gene Grossman and Michael Katz (1983, p. 752) assume the probability of conviction differs for guilty and innocent defendants, and that these probabilities are common knowledge, arguing that the prosecutor is constitutionally required to provide to the defendant all of the state's evidence against him, as well as a summary of what is necessary for conviction (*Brady v. Maryland*, 1963). By contrast, Albert Alschuler (1968, p. 66) claims that "the discovery privileges of the defense are highly restricted, and even the limited right of discovery that the law affords may be frustrated until plea negotiations are concluded."

[2] Jennifer Reinganum and Louis Wilde (1986) examined the problem of the settlement and litigation of civil suits using a signaling model. Because this paper also concerns legal bargaining with private information, it is important to distinguish this paper from the previous one. In that paper, the plaintiff had private information about the extent of damages suffered, and made a settlement demand which was either accepted or re-

the desirability of prosecutorial discretion is not resolved in complete generality, two discretionary regimes are compared. In the first, the prosecutor has discretion to offer an arbitrary sentence in exchange for a plea of guilty, with a sentence of length zero interpreted as a dismissal. In the second, plea bargaining is still permitted, but all defendants who are charged with the same crime must be offered the same sentence. It is shown that, depending upon other features of the criminal justice system and upon the preferences of society, either of these regimes may be preferred to the other. In particular, it is possible that unlimited discretion is disadvantageous for the prosecution, since it carries with it the requirement of sequential rationality.

William Landes (1971) offered the first economics-based analysis of plea bargaining. Under the assumption that the prosecutorial (and social) objective is the maximization of the sum of expected sentences subject to a resource constraint, Landes argues (among other things) that the likelihood of disposition by negotiated plea should be higher the smaller is the sentence if convicted at trial and the greater is the resource cost to the defendant of trial versus negotiated plea. Empirical analyses conducted separately at the U.S. district court level and at the county court level supported the latter implication, but yielded support for the former implication only at the U.S. district court level. William Rhodes (1976) modified the Landes model, which is based upon individual choice, to apply directly to aggregate data which are more readily available. Using data from U.S. district courts and Minnesota county courts, he found that the ratio of guilty pleas to trials is negatively correlated with the severity of the sentence offered in plea negotiations, and that increases in the defendant's resource cost of trial increase the ratio of negotiated pleas to trials. David Weimer (1978) tested a variant of the Landes model on individual case data from the Alameda County (California) Superior Court. He found that the plea-bargain sentence offer is an increasing function of the expected trial sentence (i.e., the product of the probability of conviction and the anticipated sentence upon conviction at trial), and the probability that a trial will be demanded is a decreasing function of the difference between the anticipated sentence upon conviction at trial and the sentence offered in plea negotiations. Brian Forst and Kathleen Brosi (1977) examine yet another variant of Landes' model, in which they use the length of time the prosecution carries a case as a measure of prosecutorial effort. Using data from District of Columbia courts, they find that the length of time the prosecutor will carry a case is strongly positively related to the strength of the evidence against the defendant, and mildly positively related to the seriousness of the offense (as measured by, for example, the maximum sentence applicable upon conviction at trial). While these empirical findings are all consistent with the implications of the model developed below, none of these analyses represents an appropriate structural test of this model. Thus these findings can be taken as at best weak supporting evidence.

One difficulty with Landes' theory is that the implications he tests do not strictly follow from the theory, since the actual sentence offered in a plea bargain is indeterminate (there exists a range of mutually acceptable sentence offers) if the defendant is risk averse. Moreover, in this case, the model

---

jected by the defendant. Aside from the costs of trial, the interests of the plaintiff and defendant were always opposed, since the amount of the settlement was a transfer between the parties. The issue of primary interest was how the likelihood and the amount of the settlement were affected by various litigation cost-allocation systems. In this paper, the prosecutor has private information about the strength of the case, and the defendant has private information about his guilt or innocence. The interests of the prosecutor and the defendant are not always opposed, since the prosecutor is assumed to suffer along with innocent defendants. Once equilibrium behavior has been characterized, I ask whether a particular restriction on the prosecutor's discretion in offering plea bargains can improve *ex ante* expected social welfare. Thus the nature of the private information, the forms of the objective functions, and the questions addressed by the two papers are quite different, although there are some obvious formal similarities. Previous analyses which are primarily relevant to the case of civil litigation include Lucien Bebchuk, 1984; John Gould, 1973; Ivan P'ng, 1983; Stephen Salant, 1984; William Samuelson, 1983; Steven Shavell, 1982; and Daniel Spulber, 1985.

always predicts disposition by negotiated plea; no trials should occur. In addition, the analysis presumes that all defendants are guilty, since the objective is to maximize the sum of expected sentences. Grossman and Katz (1983) admit the possibility of innocent defendants and, using an objective function which incorporates the social disutility of punishing the innocent, they provide an alternative justification of plea bargaining on the basis of its potential roles as an insurance device and as a screening device: the optimal sentence offer can induce the guilty and the innocent to self-select, so that the guilty choose negotiated pleas, while the innocent choose trial.

My analysis does not examine the desirability of the institution of plea bargaining per se; rather, it takes the existence of plea bargaining as given, and discusses the extent to which prosecutorial discretion within that framework might be desirable. It is similar to that of Grossman and Katz in its treatment of the prosecutor's (and society's) objective function. Because private information is one-sided in their model (only the defendant knows whether he is guilty or innocent), a single offer is made to all defendants; thus selective dismissals are not possible. In my model, both parties have private information (which is correlated); with unrestricted discretion, offers can be individualized and cases can be selectively dismissed. However, the regime with restricted discretion is quite similar to their model because a single sentence is offered to all defendants.

In Section I, the basic model is described and one sequential equilibrium is presented. I argue that this is an appropriate sequential equilibrium for the game with complete discretion by appealing to the work of Jeffrey Banks and Joel Sobel (1987) on "(universally) divine" sequential equilibrium and that of In-Koo Cho and David Kreps (1987) on "intuitive" sequential equilibria. In Section II, the alternative game with restricted discretion is described and its equilibrium behavior is characterized. Section III compares the welfare of the prosecutor (and, by assumption, that of society) under the two regimes, and describes circumstances under which each regime may be preferred to the other. Section IV summarizes, discusses two respects in which the basic model can be generalized, and outlines potential avenues for future research. All proofs are confined to the Appendix.

I. Unrestricted Prosecutorial Discretion

In plea bargaining, the key element need not be the guilt or innocence of the defendant (for the prosecution and the court may be unable to systematically determine the truth), but rather the *strength of the case*. In other words, the case may appear very strong, despite the fact that the defendant is innocent; similarly, the case against a guilty defendant may be very weak.

Let $t = g, i$ denote the two types of defendants: guilty and innocent, respectively. The defendant's type is assumed to be known only to the defendant. Assume that the strength of an arbitrary case is represented by the probability $\pi \in [0, 1]$ that the defendant will be found guilty at trial. The strength of the case is a summary statistic of the extent and quality of the evidence which is available to the prosecution, including verifiable information obtained from the defendant (for example, an alibi, the names of potential witnesses for the defense). Thus a defendant with $\pi = .7$ is either guilty and lucky to have a 30 percent chance of acquittal or innocent and unlucky to have a 70 percent chance of conviction. I assume that the probability of conviction is private information possessed by the prosecutor. However, the defendant's type and the prosecutor's case are not unrelated; in particular, guilt and evidence are assumed to be jointly distributed. Let $G(\pi, t)$ denote the joint distribution of π and t; that is, $G(\pi, t) =$ Prob $\{$case is of strength $\leq \pi$ and defendant is of type $t\}$. The expression which will be relevant to the prosecutor's decision is denoted $f(\pi)$ and represents the probability that, given the strength of the case, the defendant actually is guilty: $f(\pi) =$ Prob $\{$defendant is guilty$|$case is of strength $\pi\} = dG(\pi, g)/[dG(\pi, g) + dG(\pi, i)]$, where $dG(\pi, t)$ denotes the density with respect to π. Notice that $f(\pi)$ need not be the identity function; π and $f(\pi)$ can differ because different standards of evidence apply in court; for example, the prosecutor may impute a higher likelihood of guilt than conviction if in court guilt must be proved beyond a reasonable doubt. The expression which will be relevant to the defendant's decision problem is denoted $\Phi(\pi|t)$ and represents the conditional distribution of π given t: $\Phi(\pi|t) = \text{Prob}\{\text{case is of strength } \leq \pi|\text{defendant is of type } t\} = G(\pi, t)/G(1, t)$.

Although arrest itself may convey some information about the strength of the prosecutor's case (for example, if arrest requires that the evidence exceed a certain threshold value), we suppress this possibility for the moment in the interest of expositional simplicity. Similarly, one could model the discovery process as generating for the defendant a signal which is imperfectly correlated with the actual strength of the prosecution's case. These two potential information sources are discussed further in Section IV. For the present, I assume that the arrest process is essentially random, with q denoting the proportion of guilty among those arrested; that is, $q = G(1, g)$. The distribution function $G(\pi, t)$ is assumed to be common knowledge.

ASSUMPTION 1: $f'(\pi) > 0$ *for all* $\pi \in [0, 1]$; *that is, the better the case against the defendant, the greater is the likelihood of guilt.*

Define $E_t(\pi|\delta)$ to be the defendant of type t's expectation of π, given that π belongs to the set $\delta \subseteq [0, 1]$.

ASSUMPTION 2: $E_g(\pi|\delta) \geq E_i(\pi|\delta)$ *for all* δ; *that is, the distribution of* (π, t) *is such that, conditional on* $\pi \in \delta$, *a guilty defendant faces a stronger case (in expectation) than does an innocent defendant.*[3]

Grossman and Katz assumed that guilty defendants are more likely to be convicted than are innocent defendants; Assumption 2 is simply a conditional form of this assumption.

Suppose s is the sentence offered in a plea bargain, and x is the sentence anticipated upon conviction at trial; s and x are nonnegative and represented in terms of utility. The sentence upon conviction x is taken as exogenous by the prosecutor and defendant alike. Although in practice the prosecutor may be able to influence x through variations in the charges brought, I assume that sentencing following conviction at trial is fundamentally a judicial decision, which is correctly anticipated by both parties. One consequence of Assumption 2 is that, conditional on $\pi \in \delta$, a guilty defendant expects a greater punishment from trial—because he is (on average) more likely to be convicted —than his innocent counterpart. Let k denote the disutility of trial for the defendant. I have assumed that guilty and innocent defendants suffer equally in the event of trial and if given the same sentence. Grossman and Katz also assume that guilty and innocent defendants have identical utility functions. Thus my model is a natural generalization of theirs to the case in which evidence may be of variable strength. A plausible alternative assumption would be that innocent defendants suffer more in the event of trial and if given the same sentence than do guilty defendants (due to the injustice of their punishment). This possibility is considered in Section IV.

Events are assumed to proceed in the following order. First, the defendant and the prosecutor observe their private information. The prosecutor, on the basis of his private information, makes a sentence offer in exchange for a plea of guilty. Then the defendant, on the basis of his private information and whatever inference he draws from the prosecutor's offer, either accepts or rejects the offer. If he accepts the offer, both parties collect the associated payoffs. If he rejects the offer, the case must go to trial, and both parties suffer a disutility associated with trial. If the defendant is acquitted at trial, neither party receives any further payoff. Finally, if

---

[3] In the Appendix, it is shown that the distribution

$$G(\pi, i) = \eta[1 - e^{-h_i\pi}]/[1 - e^{-h_i}] \text{ and}$$

$$G(\pi, g) = (1 - \eta)[1 - e^{-h_g\pi}]/[1 - e^{-h_g}],$$

with $h_i > h_g$, satisfies Assumptions 1 and 2 for subsets δ of the form $[a, b]$.

the defendant is convicted at trial, both parties collect the payoffs associated with conviction.

Given this sequence of moves, a *strategy* for the defendant of type t is a function $p_t(s)$ specifying the probability that the defendant rejects a sentence offer of s. We can write the expected utility to a type-t defendant who is offered the sentence s and rejects it with probability ρ as

$$(1)\quad \mathrm{DU}_t(s,\rho;\delta(s)) = -\rho\left[E_t(\pi|\delta(s))x+k\right]-(1-\rho)s,$$

where $\delta(s)$ describes the defendant's *beliefs* upon observing s; that is, $\delta(s)$ is the set of prosecutor types π which the defendant believes would offer s; note that this is not subscripted, because there is no reason for different defendants to have different conjectures about this set. However, the expectation is subscripted because the innocent and the guilty may assign different distributions over the set $\delta(s)$ because (π,t) are jointly distributed. However, if the plea bargain s reveals π, then the guilty and the innocent will have identical (degenerate) expectations about π given an offer of s. In this case, they might as well use the same strategy $p(s)$.[4] This "symmetry" assumption is formalized below in Assumption 3. In the sequel I will examine only equilibria which are symmetric in this sense, but I have been unable to rule out the possibility that other "asymmetric" equilibria may exist. When the sentence offer does not reveal π, then the two types of defendants may use different strategies: $p_t(s)$, for $t=g,i$.

ASSUMPTION 3: *When both defendant types are indifferent about accepting or rejecting a sentence offer s, they use the same strategy* $p(s)$.

The objective function of the prosecutor is assumed to coincide with that of society at large, and involves three goals: appropriate punishment of the guilty, avoidance of punishment of the innocent and the conservation of resources spent on trials. The first two of these three goals are made explicit in the following excerpt from the Supreme Court opinion in *Berger v. United States* (quoted by Bruce Jackson, 1984, p. 143).

> The United States Attorney is the representative not of an ordinary party to a controversy but of a sovereignty whose obligation to govern impartially is as compelling as its obligation to govern at all; and whose interest, therefore, in a criminal prosecution is not that it shall win a case, but that justice shall be done. As such, he is in a peculiar and very definite sense the servant of the law, *the twofold aim of which is that guilt shall not escape or innocence suffer*. [Emphasis added.]

Though the last goal—that of conserving resources—is more prosaic, it is clearly of some social concern.

If c is the social cost of trial (again in terms of utility), one reasonable utility function for the prosecutor is the difference (in expected value terms) between the social valuation of the disutility of the guilty and the social valuation of the disutility of the innocent, less the social disutility of court costs. Therefore assume that society values disutility $d>0$ to a guilty defendant at γd and disutility d to an innocent defendant at $-\lambda d$, where $\gamma,\lambda>0$. A *strategy* for the prosecutor is a function $s(\pi)$ specifying the sentence offered when the case is of strength π. Then expected prosecutor utility from a case of strength π when a sentence of s is offered can be written[5,6]

---

[4] One reason to expect both types of defendant to use the same function $p(s)$ is that defense attorneys are responsible for "decoding" the sentence offer and advising the defendant about its acceptability. Since the defense attorney is also unable to verify the guilt or innocence of his client, he minimizes his own costs by describing a single function $p(s)$, rather than describing one function for innocent defendants and another for guilty defendants. Since both types of defendant will be indifferent when the equilibrium sentence offer reveals π, neither would be willing to pay more for two alternatives than for the single alternative.

$$(2)\quad \mathrm{PU}(\pi, s; p_g(s), p_i(s)) = f(\pi)\{p_g(s)[-c+\gamma(\pi x+k)] + [1-p_g(s)]\gamma s\} + (1-f(\pi)) \times \{p_i(s)[-c-\lambda(\pi x+k) - [1-p_i(s)]\lambda s\}.$$

To interpret equation (2), assume that the prosecutor has observed a case of strength π and offered a sentence of s. With probability $f(\pi)$ the defendant is actually guilty, in which case he will reject the offer with probability $p_g(s)$, forcing the case to trial, where the prosecutor suffers the disutility of trial c but gains the expected utility of punishing a guilty defendant $\gamma(\pi x+k)$. With probability $1-p_g(s)$ the defendant accepts the offer, yielding prosecutor utility of γs. With probability $1-f(\pi)$, the defendant is innocent, in which case he will reject the offer with probability $p_i(s)$, forcing the case to trial, where the prosecutor suffers the expected loss of punishing an innocent defendant $\lambda(\pi x+k)$ as well as the disutility of trial c. With probability $1-p_i(s)$ the defendant accepts the offer, yielding a utility loss of λs for the prosecutor. When both defendant types use the same strategy $p(s)$, this simplifies to

$$(3)\quad \mathrm{PU}(\pi, s; p(s)) = p(s)[-c+a(\pi)(\pi x+k)] + [1-p(s)]a(\pi)s,$$

where $a(\pi)=f(\pi)\gamma-(1-f(\pi))\lambda$. The expression $a(\pi)$ represents the expected net social utility of an additional unit of disutility imposed upon a defendant against whom there is a case of strength π, where the expectation arises from the fact that the defendant's guilt is unverifiable. Notice that $a'(\pi)=f'(\pi)(\gamma+\lambda)>0$. Define π_0 such that $a(\pi_0)=0$. If $\pi_0 \in (0,1)$ exists, it will be unique, and hereafter we assume its existence and interiority. Also define $A(s)=\int a((s-k)/x)ds$; note that $A(\cdot)$ is an increasing function for $s>\pi_0 x+k$.

For a general definition of sequential equilibrium, see David Kreps and Robert Wilson (1982); for our purposes, the following definition will suffice.

Definition 1: A *sequential equilibrium* consists of beliefs $\delta^*(\cdot)$ and strategies $(p_g^*(\cdot), p_i^*(\cdot), s^*(\cdot))$ such that

(a) $p_t^*(s)$ maximizes $\mathrm{DU}_t\ (s, p; \delta^*(s))$, for $t=g,i$;

(b) $s^*(\pi)$ maximizes $\mathrm{PU}(\pi,\ s;\ p_g^*(s), p_i^*(s))$; and

(c) $\delta^*(s)\subseteq[0,1]$ for all s, and $\delta^*(s)=\{\pi \mid s=s^*(\pi)\}$ whenever this set is nonempty.

That is, the equilibrium strategy of each defendant type maximizes that defendant's expected utility, given the beliefs. The prosecutor's strategy maximizes the prosecutor's expected utility, given the anticipated responses of the two types of defendants. Finally, the defendants' beliefs are always con-

---

[5] This specification should be regarded as a *local* utility function, assumed to be valid at least up to the maximum penalty $x+k$. Since society has chosen x as the penalty given conviction, and the trial process itself imposes k, it must be assumed that society believes $x+k$ to be an appropriate penalty. Thus it is plausible to assume that social utility increases with the penalty to the guilty as least until $x+k$. Since no defendant would accept a plea bargain which specified a sentence greater than $x+k$, we need not specify social utility for such penalties. But this specification is not inconsistent with the assumption that social utility declines when penalties greater than $x+k$ are imposed on guilty defendants.

[6] One vocal opponent of the practice of plea bargaining (Albert Alschuler, 1981, p. 710) would reject this as an appropriate social objective function on several accounts. One is that it relies upon the prosecutor's estimate of guilt, $f(\pi)$, rather than upon an actual trial outcome. "Society should not legitimate the prosecutor's presumption that he can determine guilt more accurately than a court or jury..." Another objection is simply that the administration of justice is not to be traded off against the conservation of resources. I do not deal with this issue here; obviously the formulation of an appropriate social objective is an important issue, but it goes well beyond the scope of this paper.

fined to the set of prosecutor types which are known to be possible, and these beliefs are correct for equilibrium sentence offers.

PROPOSITION 1: *A sequential equilibrium for this model is for the prosecutor to offer* $s^* = 0$ (i.e., *dismiss the case*), *if* $\pi < \pi_0$; *otherwise offer* $s^* = \pi x + k$. *Let* $\underline{s} = \pi_0 x + k$ *and* $\bar{s} = x + k$. *Then the defendant* (*whether guilty or innocent*) *rejects the offer* s *with probability* $p^*(s) = 1$ *if* $s > \bar{s}$, *with probability* $p^*(s) = 1 - \exp\{[A(\underline{s}) - A(s)]/c\}$ *for* $s \in [\underline{s}, \bar{s}]$, *and with probability* $p^*(s) = 0$ *if* $s < \underline{s}$. *Finally, the defendants' beliefs are* $\delta^*(s) = 1$ *for* $s \geq \bar{s}, \delta^*(s) = (s-k)/x$ *for* $s \in [\underline{s}, \bar{s}]$, $\delta^*(s) = \pi_0$ *for* $s \in (0, \underline{s})$, *and* $\delta^*(s) = [0, \pi_0)$ *for* $s = 0$.

Notice that sufficiently weak cases—those with $\pi < \pi_0$—are dismissed. How weak is "sufficiently weak" depends only upon the form of the inference function $f(\pi)$ and upon $\lambda/(\gamma + \lambda)$, and not on the resource cost of trials. That is, a case is never dismissed on account of the resource costs of pursuing it; this decision is based only on the merits of the case. If equal consideration is given to punishing the guilty and avoiding punishing the innocent (i.e., $\lambda = \gamma$), then to be dismissed a case must leave the prosecutor believing innocence is more likely than guilt ($f(\pi) < 1/2$). If society is more concerned with avoiding the punishment of the innocent than with ensuring the punishment of the guilty (i.e., $\lambda > \gamma$), then the prosecutor will dismiss some cases in which guilt is more likely than innocence. The result that the prosecution will dismiss cases in which there is sufficient doubt regarding the defendant's guilt is consistent with anecdotal evidence. Alschuler (1981, p. 708) remarks that "...when a prosecutor does entertain serious doubts concerning a defendant's factual guilt, he is likely to decline to prosecute..." and Charles Silberman (1980, p. 367) concludes that "Most prosecutors believe that they should not press charges unless they are convinced of the defendant's guilt."

Formally, the equilibrium consists of two portions. One involves complete pooling for prosecutor types $\pi < \pi_0$ while the other involves complete separation for types $\pi \geq \pi_0$. Under the assumption that when the sentence offer reveals the strength of the case both types of defendant use the same strategy $p(s)$, the uniqueness proof of Reinganum and Wilde (1986) can be adapted to show that the separating portion of the equilibrium is unique. If this symmetry assumption is relaxed, I have been unable to rule out the possibility of a separating equilibrium in which the two defendant types use different equilibrium strategies.[7]

Signaling models typically possess a multiplicity of equilibria, some of which depend upon the specification of out-of-equilibrium beliefs. In view of the potential sensitivity of equilibrium behavior to out-of-equilibrium beliefs, it is worthwhile examining alternatives to those beliefs specified in Proposition 1. In the Appendix, it is shown that one sufficient condition for the equilibrium strategies specified in Proposition 1 to be supported by *any* out-of-equilibrium beliefs is the parametric restriction $\pi_0 \leq c/\lambda x$. Alternatively, if $\delta^*(s)$ is a singleton for $s \in (0, \underline{s})$, then any such beliefs will support the equilibrium behavior specified in Proposition 1. Thus under plausible circumstances this equilibrium behavior is "(universally) divine" (Banks and Sobel, 1987) and "intuitive" (Cho and Kreps, 1987). Indeed, it will survive any equilibrium refinement which relies only on belief restrictions. Although this does not imply that this is the unique such equilibrium, it does suggest that it is a highly plausible one which deserves consideration and close inspection.

---

[7] However, some likely candidates can be ruled out. For instance, a candidate in which the innocent defendant rejects s when indifferent while a guilty defendant randomizes would be characterized by $\tilde{s}(\pi) = \pi x + k$ for $\pi \in [0,1]$. Thus $\underline{s} = k$ and $\bar{s} = x + k$. Strategies for the defendants: $\tilde{p}_i(s) = \tilde{p}_g(s) = 1$ for $s > \bar{s}, \tilde{p}_i(s) = 1$ and $\tilde{p}_g(s) = 1 - \exp\{\gamma(\underline{s} - s)/c\}$ for $s \in [\underline{s}, \bar{s}]$ and $\tilde{p}_i(s) = \tilde{p}_g(s) = 0$ for $s \in [0, \underline{s})$ are supported by the beliefs $\delta(s) = (s-k)/x$ for $s \in [\underline{s}, \bar{s}]$ and for arbitrary out-of-equilibrium beliefs $\tilde{\delta}(s) \subseteq [0,1]$. Although $\tilde{s}(\pi)$ provides a stationary point of $\mathrm{PU}(\pi, s; \tilde{p}_g(s), \tilde{p}_i(s))$, it does not provide a maximum for π sufficiently near 0. To see this, note that $\mathrm{PU}(\pi, \tilde{s}(\pi); \tilde{p}_g(\tilde{s}(\pi)), \tilde{p}_i(\tilde{s}(\pi))) = a(\pi)\tilde{s}(\pi) - [f(\pi)\tilde{p}_g(\tilde{s}(\pi)) + (1 - f(\pi))]c$, while $\mathrm{PU}(\pi, 0; \tilde{p}_g(0), \tilde{p}_i(0)) = 0$. Evaluating both expressions at $\pi = 0$ implies that $\mathrm{PU}(0, 0; \tilde{p}_g(0), \tilde{p}_i(0)) > \mathrm{PU}(0, \tilde{s}(0); \tilde{p}_g(\tilde{s}(0)), \tilde{p}_i(\tilde{s}(0)))$, if and only if $0 > a(0)\underline{s} - [1 - f(0)]c$. Since $a(0) < 0$, the inequality holds. Thus prosecutors with sufficiently weak cases (i.e., π sufficiently near 0) would prefer $s = 0$ to $\tilde{s}(\pi)$.

Another expression of interest is the equilibrium probability of trial as a function of the strength of the case. This is the composition of the equilibrium probability of rejection and the equilibrium sentence offer: $\hat{p}(\pi)=p^*(s^*(\pi))$.

$$\hat{p}(\pi)=1-\exp\{[A(\underline{s})-A(s^*(\pi))]/c\}. \tag{4}$$

PROPOSITION 2: (*a*) *For* $s\in(\underline{s},\bar{s})$, *the equilibrium probability of rejection* $p^*(s)$ *increases with an increase in the sentence* s *offered in exchange for a guilty plea, and the social weight* γ *given to the guilty defendant's disutility*; $p^*(s)$ *decreases with an increase in the social cost of trial* c, *the defendant's disutility of trial* k, *the sentence anticipated upon conviction at trial* x, *and the social weight* λ *given to the innocent defendant's disutility.*

(*b*) *For* $\pi\in[\pi_0,1]$, *the equilibrium sentence* $s^*(\pi)$ *offered in exchange for a plea of guilty increases with an increase in the strength of the case* π, *the sentence anticipated upon conviction at trial* x, *and the defendant's disutility of trial* k. *It is unaffected by the parameters* c, γ, *and* λ.

(*c*) *For* $\pi\in(\pi_0,1]$, *the equilibrium probability of trial* $\hat{p}(\pi)$ *increases with the strength of the case* π, *and the social weight* γ *given to the guilty defendant's disutility*; $\hat{p}(\pi)$ *decreases with an increase in the social cost of trial* c, *and the social weight* λ *given to the innocent defendant's disutility. The effects of* k *and* x *upon* $\hat{p}(\pi)$ *are indeterminate because there are two conflicting effects*; *the direct effect is to reduce the probability that a given sentence offer* s *is rejected*, *but the indirect effect is to raise the equilibrium sentence offer* s^*, *which in turn increases the likelihood of rejection.*

An important implication of equilibrium is that when a case is not dismissed, the likelihood that is will be resolved by a guilty plea is greater the weaker is the case. This is one plausible interpretation of Alschuler's (1968, p. 60) statement that "...the greatest pressures to plead guilty are brought to bear on defendants who may be innocent. The universal rule is that the sentence differential between guilty-plea and trial defendants increases in direct proportion to the likelihood of acquittal," where the "sentence differential" is defined as the difference between the sentence upon conviction x and the realized sentence $s^*(\pi)$. The reader should note that I quote Alschuler's statement in order to invoke its empirical claim, without necessarily endorsing his terminology or his interpretation of plea bargaining as the use of pressure tactics. As I remarked in the introductory discussion, the comparative statics implications of this model are also consistent with the empirical findings of Forst and Brosi, 1977; Landes, 1971; Rhodes, 1976; and Weimer, 1978.

II. Restricted Prosecutorial Discretion

The model of Section I involved considerable discretion upon the part of the prosecutor. It has been argued that such discretion is undesirable because it gives rise to horizontal inequities: defendants charged with the same crime and subject to the same penalties upon conviction are offered different sentences in plea negotiations. As Alschuler (1968, p. 60) puts it, "the practice of bargaining hardest when the case is weakest leads to grossly disparate treatment for identical offenders" (where by "bargaining hardest" he means offering the greatest sentence differential between trial and negotiated plea). Such "inequities" are eliminated if prosecutors are constrained to offer the same plea bargain to all such defendants. Economists and some legal scholars (for example, Thomas Church, 1979) are likely to reject this notion of horizontal inequity by arguing that defendants who face a different probability of conviction are indeed "different" defendants, or by invoking an alternative notion of *ex ante* horizontal equity. While this controversy is lively and interesting, in Section III, I examine an alternative rationale for such a constraint; I ask whether it can improve the prosecutor's (and society's) *ex ante* expected utility.[8] In order to answer this question, it is first necessary to characterize equilibrium behavior in a regime of restricted discretion.

When the prosecutor is required to offer the same sentence to every defendant who is charged with the same crime, independent of the strength of the prosecution's case, he must make a "pooling" offer. This can result in self-selection by the guilty and innocent defendants, because they have different *ex ante* expected values of π. Let E_t denote the defendant of type t's prior expectation over π:

$$E_t = \int_0^1 \pi d\Phi(\pi | t).$$

By Assumption 2, $E_g \geq E_i$; the guilty defendant expects a greater likelihood of conviction than does the innocent defendant.

If a pooled offer s is made, it is the expectation E_t which governs the defendant's decision.

$$\mathrm{DU}_t(s, \rho; [0,1]) = -\rho(E_t x + k) - (1-\rho)s.$$

The innocent defendant rejects s if and only if $s > s_0 = E_i x + k$, while the guilty defendant rejects s if and only if $s > s^0 = E_g x + k$. Thus any offer $s \in (s_0, s^0]$ will be rejected by the innocent and accepted by the guilty.

The prosecutor must determine an optimal sentence offer, given the anticipated behavior of the defendant types. If an offer of $s \in (s_0, s^0]$ is made, it is accepted by the guilty and rejected by the innocent; thus the best such offer is $s = s^0$, which yields *ex ante* expected prosecutor utility

$$(5) \quad U_1 = q\gamma s^0 - (1-q)[c + \lambda(E_i x + k)].$$

Any offer $s > s^0$ is rejected by all defendants, yielding *ex ante* expected prosecutor utility

$$(6) \quad U_2 = -c + q\gamma(E_g x + k) - (1-q)\lambda(E_i x + k).$$

Finally, an offer $s \leq s_0$ is accepted by all defendants, yielding *ex ante* expected prosecutor utility

$$(7) \quad U_3(s) = [q\gamma - (1-q)\lambda]s.$$

When $q\gamma - (1-q)\lambda > 0$, the optimal such offer is s_0; when $q\gamma - (1-q)\lambda < 0$, the optimal such offer is $s = 0$ (i.e., dismiss all cases). When $q\gamma - (1-q)\lambda = 0$, any offer is $[0, s_0]$ is optimal. A comparison of equations (5), (6), and (7) yields the following characterization of equilibrium with restricted discretion.

PROPOSITION 3: (*A*) *Suppose that* $\lambda x(E_g - E_i) \leq c$. *Define three intervals*:

$$I_{1A} = [0, \lambda/(\lambda+\gamma)];$$

$$I_{2A} = [\lambda/(\lambda+\gamma), c/(\gamma x(E_g - E_i) + c)];$$

and $I_{3A} = [c/(\gamma x(E_g - E_i) + c), 1]$. *The optimal sentence offer with restricted discretion is* $s = 0$ *for* $q \in I_{1A}$; *that is, all cases are dismissed. The optimal offer is* $s = s_0$ *for* $q \in I_{2A}$; *that is, all cases are settled by negotiated plea at a sentence of* s_0. *Finally, the optimal sentence offer is* $s = s^0$ *for* $q \in I_{3A}$; *that is, guilty defendants plead guilty in exchange for a sentence of* s^0, *while innocent defendants go to trial. Expected ex ante prosecutor utility with restricted discretion is given by*

$$(8) \quad \mathrm{EPU}^r = \begin{cases} 0 & q \in I_{1A}, \\ [q\gamma - (1-q)\lambda](E_i x + k) & q \in I_{2A}, \\ q\gamma(E_g x + k) - (1-q) & \\ \times [c + \lambda(E_i x + k)] & q \in I_{3A}. \end{cases}$$

(*B*) *Suppose* $\lambda x(E_g - E_i) \geq c$. *Define two intervals*: $[I_{1B} = [0, (c + \lambda(E_i x + k))/$

---

[8] Externally imposed constraints upon a player's strategy set can enable him to make binding commitments, which can improve his payoff in sequential equilibrium. What is of interest is whether this particular constraint can be welfare-improving. An externally imposed constraint serves to coordinate beliefs and makes a pooling offer credible in the face of equilibrium refinements such as universal divinity.

$(c+\lambda(E_i x+k)+\gamma(E_g x+k))]$; *and*]

$$I_{2B}=\left[(c+\lambda(E_i x+k))/\right.$$

$$\left.(c+\lambda(E_i x+k)+\gamma(E_g x+k)),1\right].$$

The optimal sentence offer with restricted discretion is $s=0$ *for* $q\in I_{1B}$; *that is, all cases are dismissed. The optimal offer is* s^0 *for* $q\in I_{2B}$; *that is, guilty defendants plead guilty in exchange for a sentence of* s^0, *while innocent defendants go to trial. Ex ante expected prosecutor utility with restricted discretion is given by*

$$(9)\qquad \mathrm{EPU}^r=\begin{cases}0 & q\in I_{1B},\\ q\gamma(E_g x+k)-(1-q) & \\ \times[c+\lambda(E_i x+k)] & q\in I_{2B}.\end{cases}$$

It is interesting to note that when the prosecutor types separate (that is, the offer is a revealing one), the defendant types pool, but when the prosecutor types pool, the defendant types may separate (i.e., the guilty plead at s^0 and the innocent go to trial). The screening outcome is much the same as that of Grossman and Katz (1983), and suffers from the same disturbing feature; the prosecutor knows that each case he takes to trial is against an innocent defendant.

III. Characterization of the Preferred Regime

In this section, we wish to compare equilibrium *ex ante* expected prosecutor utility in the regimes of unrestricted and restricted discretion described in Sections I and II, respectively. A convenient characterization of the circumstances under which discretion is preferred can be made in terms of the exogenous expression q.

Define

$$\mathrm{PU}^*(\pi)=\mathrm{PU}(\pi,s^*(\pi);\,p^*(s^*(\pi))).$$

By the envelope theorem,

$$\begin{aligned}\mathrm{PU}^{*\prime}(\pi)&=\partial\,\mathrm{PU}(\pi,s(\pi);\,p^*(s^*(\pi)))/\partial\pi\\ &=a'(\pi)(\pi x+k)\\ &\quad+p^*(s^*(\pi))\,a(\pi)x>0\end{aligned}$$

for $\pi\geq\pi_0$. Thus $\mathrm{PU}^*(\pi)>0$ for $\pi>\pi_0$, while $\mathrm{PU}^*(\pi)=0$ for $\pi\leq\pi_0$. The density function for π can be written $qd\Phi(\pi|g)+(1-q)d\Phi(\pi|i)$. Thus the prosecutor's equilibrium *ex ante* expected utility with discretion can be written

$$(10)\qquad \mathrm{EPU}^*=\int_0^1\mathrm{PU}^*(\pi)\,qd\Phi(\pi|g)+\int_0^1\mathrm{PU}^*(\pi)(1-q)\,d\Phi(\pi|i).$$

Note that $\mathrm{EPU}^*>0$ since $\mathrm{PU}^*(\pi)>0$ for $\pi>\pi_0$ and $\mathrm{PU}^*(\pi)=0$ for $\pi\leq\pi_0$. The expression EPU* can be compared with *ex ante* expected prosecutor utility under restricted discretion EPU^r to obtain the following result.

PROPOSITION 4: *For given values of the parameters* $E_i, E_g, \gamma, \lambda, c, k$, *and* x, *there exists a unique* $q_0\in(0,1)$ *such that unrestricted discretion is preferred for* $q<q_0$ *and restricted discretion is preferred for* $q>q_0$.

Intuitively, when the arrest process does not do a good job of screening out the innocent (i.e., $q<q_0$), then discretion at the prosecution stage is preferred. On the other hand, when the prosecutor can be confident that most defendants are guilty, the restriction to a uniform offer is preferred. Many factors influence the comparison between discretionary regimes. Recall that in the regime with restricted discretion, all defendants accept sentence offers at or below $s_0=E_i x+k$, while only guilty defendants accept sentence offers in the range $(s_0,s^0]$, where $s^0=E_g x+k$. With a sufficiently high proportion of innocent defendants the prosecutor optimally dismisses all cases; an offer in $(0,s_0)$ is too costly in terms of punishing the innocent, while an offer in $(s_0,s^0]$ is too costly in terms of the trials generated by innocent defendants' rejection of the plea offer. In the regime with unrestricted discretion the prosecutor can selectively dismiss cases while retaining the ability to impose penalties upon defendants who are likely to be guilty. Thus discretion is preferred. With a sufficiently high proportion of guilty defendants, the optimal sentence offer with re-

stricted discretion is s^0; since innocent defendants reject this offer, it is never imposed upon them. Since there are relatively few innocent defendants, the costs generated by their trials and occasional erroneous convictions are not great. With unrestricted discretion, there is always a fraction of cases involving guilty defendants which goes to trial; these trials generate court costs and occasional erroneous acquittals. Thus restricted discretion is preferred.

It is straightforward to determine which discretionary regime is preferred by each type of defendant by comparing $\mathrm{EDU}_t^* = -\int_{\pi_0}^1(\pi x + k)d\Phi(\pi|t)$ with

$$\mathrm{EDU}_t^r = \begin{cases} 0 & q \in I_{1A} \text{ or} \\ & \quad q \in I_{1B}, \\ -\int_0^1(\pi x + k)d\Phi(\pi|i) & q \in I_{2A}, \\ -\int_0^1(\pi x + k)d\Phi(\pi|t) & q \in I_{3A} \text{ or} \\ & \quad q \in I_{2B}, \end{cases}$$

for $t = g, i$, where the sets I_{jk} are those defined in Proposition 3. For $q \in I_{1A}$ or I_{1B}, both innocent and guilty defendants prefer unrestricted discretion; for $q \in I_{3A}$ or I_{2B}, both prefer restricted discretion; finally, for $q \in I_{2A}$, either both prefer unrestricted discretion (for π_0 large) or innocent defendants prefer unrestricted discretion while guilty defendants prefer restricted discretion (for π_0 small).

These results can be combined with those of Proposition 4 to show that each of the following preference patterns arises from some configuration of parameters: (1) both types of defendants prefer unrestricted discretion, while the prosecutor prefers restricted discretion; (2) both types of defendants prefer restricted discretion while the prosecutor prefers unrestricted discretion; (3) all parties unanimously prefer unrestricted discretion; (4) innocent defendants and the prosecutor prefer unrestricted discretion while guilty defendants prefer restricted discretion; and (5) guilty defendants and the prosecutor prefer restricted discretion while innocent defendants prefer unrestricted discretion. For most parameter values, there will be some disagreement among the interested parties regarding the preferred discretionary regime, so the existence of controversy is not surprising. When unanimity does occur, it favors the regime of unrestricted discretion.

IV. Conclusions

The essential features of plea-bargaining equilibrium which emerged from this analysis are that sufficiently weak cases are dismissed, where this sufficiency does not depend upon the resource cost of trial but upon the social costs and benefits of punishing the innocent and the guilty, respectively; that defendants against whom a sufficiently strong case exists are offered a sentence (in exchange for a plea of guilty) which increases with the likelihood of conviction at trial and the defendant's anticipated disutility of trial and conviction; and finally, the defendants are more likely to reject higher sentence offers, so that the likelihood of trial is an increasing function of the strength of the case.

A uniform-offer restriction upon prosecutorial discretion was found to improve *ex ante* expected welfare when the proportion of guilty among those arrested is sufficiently high. Typically the prosecutor prefers the disposition of cases involving guilty defendants under restricted discretion, and prefers the disposition of cases involving innocent defendants under unrestricted discretion (see equation (A11) in the Appendix). When the proportion of guilty among those arrested is sufficiently high, the benefits from restricted discretion outweigh the costs.

Thus far I have assumed that arrest was a totally random process. However, arrest itself may convey information about the prosecutor's case if arrest requires that the evidence exceed a certain standard. For instance, if arrest requires that the case be of strength at least π_α, then the defendant of type t can infer that the relevant density function for π is $d\Phi(\pi|t)/[1-\Phi(\pi_\alpha|t)]$. In this case, the proportion of guilty among those arrested is

$$q(\pi_\alpha) = \int_{\pi_\alpha}^1 dG(\pi, g) \Big/ \left[\int_{\pi_\alpha}^1 dG(\pi, g) + \int_{\pi_\alpha}^1 dG(\pi, i)\right].$$

It can be shown that, for the truncated exponential distribution mentioned in Section II and examined in the Appendix, $q'(\pi_\alpha) > 0$; that is, a higher arrest standard increases the proportion of guilty among those arrested. Although their statements need to be modified (in the obvious ways), Propositions 1–4 also hold with this specification of the model. This suggests that arrest standards are important determinants of the institutional forms of related aspects of the criminal justice system. One effect of low arrest standards is to make prosecutorial discretion more attractive since low arrest standards generate a higher fraction of cases likely to involve innocents, cases which must be weeded out at the prosecution stage. On the other hand, high arrest standards make restricted discretion more attractive. Of course, the arrest standard also has a direct effect upon *ex ante* expected social utility, since it determines the number of cases as well as the fraction likely to involve a guilty defendant. The determination of the optimal arrest standard and the associated discretionary regime is beyond the scope of this paper, but some of the relevant tradeoffs have been identified. A more difficult task is to incorporate the discovery process. One way to do this is to assume that the defendant receives a signal which is (imperfectly) correlated with the actual strength of the case. If the prosecutor also observes this signal, then this is basically an exercise in updating priors. Both parties simply use their posterior distributions in their decision making. If the signal is private information for the defendant, matters could become considerably more complicated.

One might plausibly argue that innocent defendants are likely to suffer more than guilty defendants in the same circumstances because their punishment is undeserved. One natural way to model this would be to assume that a sentence yielding disutility d to a guilty defendant yields disutility αd to an innocent defendant, where $\alpha > 1$. This "injustice" multiplier should be present whether the sentence is imposed upon conviction at trial or upon the acceptance of a plea bargain. Since the disutility of trial is also unjustly imposed upon innocent defendants, a trial yielding disutility k to a guilty defendant yields disutility αk to an innocent defendant. Under these reinterpretations, the innocent defendant who accepts a plea bargain of s receives disutility αs, while one who chooses trial expects $E_i(\pi|\delta(s))\alpha x + \alpha k$. Thus the innocent defendant's expected utility is now simply scaled up by the factor α. Since it is not the absolute level of disutility which matters, but the comparison between the disutility of going to trial and that of accepting a plea bargain, the innocent defendant's decision problem is unchanged. The prosecutor's objective function has the same functional form, but the parameter λ is now replaced by $\alpha\lambda$. Subject to this substitution, the equilibrium strategies are as given in Proposition 1. Of course, justice might enter the utility function in other than a multiplicative manner, in which case innocent and guilty defendants will likely behave differently in equilibrium.

These results have been obtained in a very streamlined model. Even now, several issues remain open. First, there is the possibility of asymmetric separating equilibria in which both types of defendants randomize, but according to different strategies. In addition, there are likely to be equilibria which involve some degree of pooling. I have taken a broad approach to equilibrium refinement, presenting sufficient conditions for the proposed equilibrium to be robust to arbitrary out-of-equilibrium beliefs. However, the application of specific refinement concepts will imply particular restrictions on out-of-equilibrium beliefs. Due to the difficulty of applying these concepts to the case of two-sided asymmetric information, this has been left for future research.

Finally, many discussions of the desirability or deplorability of prosecutorial discretion center on alleged problems of a principal-agent nature. Thus it would be interesting to compare these results with those which might emerge under alternative prosecutorial objective functions. For example, one common assertion is that prosecutors care more about their "winning percentage" at trial than about public welfare, the implication being that they are willing to try only cases they are likely to win. But the results of this paper imply that stronger cases are more likely to go to trial (in equilibrium) even absent such a divergence between the preferences of society and the prosecutor.

Some conclusions about the likely effects of divergent objectives can be obtained by assuming that the prosecutor's objective has the same form as that of society, but that he misperceives the parameters. For instance, if he overestimates the social value of punishing the guilty (or underestimates the social value of protecting the innocent from punishment), he will pursue cases he should have dismissed, and the likelihood of trial in each case will be too high. On the other hand, if the prosecutor overestimates the social resource cost of trial, the likelihood of trial in each case will be too low. While these implications are suggestive, a more thorough investigation of alternative prosecutorial objective functions is necessary in order to make any inferences about objectives on the basis of observed behavior.

APPENDIX

Example 1: Let $G(\pi, i) = \eta[1 - e^{-h_i\pi}]/[1 - e^{-h_i}]$ and $G(\pi, g) = (1-\eta)[1 - e^{-h_g\pi}]/[1 - e^{-h_g}]$, where $h_i > h_g$ and $\pi \in [0,1]$. We wish to verify that Assumptions 1 and 2 (for δ of the form [a, b]) are satisfied.

ASSUMPTION 1: *$f'(\pi) > 0$ for $\pi \in [0,1]$.* The probability of guilt conditional upon π is $f(\pi) = dG(\pi, g)/[dG(\pi, g) + dG(\pi, i)]$. Differentiating and collecting terms implies that $f'(\pi) > 0$ if and only if $dG(\pi, i)/d^2G(\pi, i) > dG(\pi, g)/d^2G(\pi, g)$. For our example, this inequality becomes $-1/h_i > -1/h_g$ or $h_i > h_g$.

ASSUMPTION 2: *(for δ of the form $[a, b]$).* $E_g(\pi|\delta) \geq E_i(\pi|\delta)$ for all δ. The conditional expectation of π, given the defendant's type and given that $\pi \in [a, b]$ is

$$E_t(\pi\|[a, b]) = \int_a^b \pi d\Phi(\pi|t)/[\Phi(b|t) - \Phi(a|t)]$$

$$= a + \int_a^b \{[\Phi(b|t) - \Phi(\pi|t)]/$$

$$[\Phi(b|t) - \Phi(a|t)]\}\, d\pi.$$

For our example, $[\Phi(b|t) - \Phi(\pi|t)]/[\Phi(b|t) - \Phi(a|t)] = [e^{-h_t\pi} - e^{-h_tb}]/[e^{-h_ta} - e^{-h_tb}]$. Since $w(h) = [e^{-h\pi} - e^{-hb}]/[e^{-ha} - e^{-hb}]$ is nonincreasing in h for $\pi \in [a, b]$, $h_i > h_g$ implies $w(h_i) \leq w(h_g)$ for $\pi \in [a, b]$. Thus $E_i(\pi\|[a, b]) \leq E_g(\pi\|[a, b])$.

Proof of Proposition 1. Given $\delta^*(s)$, is $p^*(s)$ optimal? For $s > \bar{s}, \delta^*(s) = 1$ and $\mathrm{DU}_t = -\rho(x+k) - (1-\rho)s$, so $p^*(s) = 1$ is optimal. For $s \in [\underline{s}, \bar{s}], \delta^*(s) = (s-k)/x$ and $\mathrm{DU}_t = -\rho s - (1-\rho)s$, so any ρ works. Thus $p^*(s) = 1 - \exp\{[A(\underline{s}) - A(s)]/c\}$ is optimal for $s \in [s, \underline{s}]$. For $s \in (0, \underline{s}), \delta^*(s) = \pi_0$ and $\mathrm{DU}_t = -\rho(\pi_0 x + k) - (1-\rho)s$, so $p^*(s) = 0$ is optimal for $s \in (0, \underline{s})$. Finally, for $s = 0, \delta^*(s) = [0, \pi_0)$; since $E_t(\pi\|[0, \pi_0)) \geq 0, \mathrm{DU}_t = -\rho[E_t(\pi\|[0, \pi_0))x + k] - (1-\rho)s$ implies that $p^*(s) = 0$ is optimal.

Given $p^*(s)$, is $s^*(\pi)$ optimal? Since $p^*(s)$ is the same for both defendant types, we can use the prosecutor's payoff as described in equation (3). For $\pi < \pi_0$, $\mathrm{PU}(\pi, s; p^*(s)) < 0$ for all positive values of s while $\mathrm{PU}(\pi, 0; p^*(0)) = 0$, so $s^*(\pi) = 0$ is optimal for $\pi < \pi_0$. For $\pi \geq \pi_0, \mathrm{PU}(\pi, \bar{s}; p^*(\bar{s})) > \mathrm{PU}(\pi, s; p^*(s))$ for $s > \bar{s}$ and $\mathrm{PU}(\pi, \underline{s}; p^*(\underline{s})) \geq \mathrm{PU}(\pi, s; p^*(s))$ for $s < \underline{s}$ (with equality only at $\pi = \pi_0$). So $s^*(\pi) \in [\underline{s}, \bar{s}]$ for $\pi \geq \pi_0$. Differentiating $\mathrm{PU}(\pi, s; p^*(s))$ with respect to s and equating to zero yields

$$d\mathrm{PU}/ds = \exp\{[A(\underline{s}) - A(s)]/c\}$$

$$\times \{[A'(s)/c][-c + a(\pi)(\pi x + k - s)]$$

$$+ a(\pi)\} = 0.$$

Upon noting that $\exp\{\cdot\}$ is never zero, and that $A'(s) = a((s-k)/x)$, this simplifies to

$$a(\pi)(\pi x + k - s) = c[a((s-k)/x) - a(\pi)].$$

Since $a(\cdot)$ is an increasing function, whenever the left-hand side is positive, the right-hand side is negative, and vice versa. Thus the only solution is $s = \pi x + k$. That is, the function $\mathrm{PU}(\pi, s; p^*(s))$ has a unique stationary point at $s^* = \pi x + k$. Moreover, the second derivative of $\mathrm{PU}(\pi, s; p^*(s))$ with respect to s is negative at s^*. Thus s^* provides a local maximum. But it is also a global maximum because otherwise there would have to be an interior local minimum between the local maxima, but there are no other stationary points of $\mathrm{PU}(\pi, s, p^*(s))$.

Finally, we need to check consistency: $\delta^*(s) \subseteq [0,1]$ for all s. Moreover, $\delta^*(s^*(\pi)) = \pi$ for $\pi \in [\pi_0, 1]$ and $\delta^*(0) = [0, \pi_0)$, which is exactly the set of cases dismissed by the prosecutor (i.e., offered $s^*(\pi) = 0$).

Alternative Out-of-Equilibrium Beliefs. I claim that a sufficient condition for the equilibrium strategies of Proposition 1 to be robust to out-of-equilibrium beliefs is: $\pi_0 \leq c/\lambda x$. To see this, recall that since $E_g(\pi|\delta(s)) \geq E_i(\pi|\delta(s))$ for all $\delta(s)$ (by Assumption 2), if an innocent defendant strictly prefers to accept s, then so does a guilty defendant. Thus (assuming identical behavior when both defendant types are indifferent), only three types of asymmetric behavior can arise: (1) g accepts s and i rejects s; (2) g randomizes due to indifference and i rejects s; and (3) g accepts s and i randomizes due to indifference.

Consider first the case of identical behavior. For $s \notin 0 \cup [\underline{s}, \bar{s}]$,

$$\mathrm{PU}(\pi, s; p(s)) = p(s)[-c + a(\pi)(\pi x + k)]$$

$$+ [1 - p(s)]a(\pi)s$$

for some $p(s) \in [0,1]$. Note that any $s > \bar{s}$ is optimally rejected by both defendant types regardless of their beliefs. Let $PU^*(\pi) = PU(\pi, s^*(\pi); p^*(s^*(\pi)))$. Since $PU^*(\pi) = a(\pi)(\pi x + k) - p^*(s^*(\pi))c > a(\pi)(\pi x + k) - c$ for all $\pi \geq \pi_0$, no such π would offer $s > \bar{s}$. Similarly, $PU^*(\pi) = 0 > a(\pi)(\pi x + k) - c$ for $\pi < \pi_0$, so no such π would offer $s > \bar{s}$.

For $s \in (0, \underline{s})$, when the same behavior $p(s)$ arises, $PU(\pi, s; p(s)) < 0 = PU^*(\pi)$ for $\pi < \pi_0$. For $\pi \geq \pi_0$, notice that

$$PU(\pi, s; p(s)) \leq p(s)[-c + a(\pi)(\pi x + k)] + [1 - p(s)]a(\pi)(\pi_0 x + k).$$

But $PU^*(\pi) = a(\pi)(\pi x + k) - p^*(s^*(\pi))c > a(\pi)(\pi x + k) - c$ and $PU^*(\pi) \geq PU(\pi, \underline{s}, p^*(\underline{s})) = a(\pi)(\pi_0 x + k)$ for all $\pi \geq \pi_0$. Thus no $\pi \geq \pi_0$ would prefer $s \in (0, \underline{s})$ to $s^*(\pi)$.

Recall that whenever $\delta^*(s)$ is a singleton, both defendant types have identical expectations and are assumed to behave in an identical manner; thus the equilibrium behavior of Proposition 1 is robust to any out-of-equilibrium beliefs which consist only of singletons.

Next consider asymmetric behavior of type (1): g accepts s and i rejects s, for $s \in (0, \underline{s})$. Then

$$\begin{aligned} PU(\pi, s; 0, 1) &= f(\pi)\gamma s + [1 - f(\pi)][-c - \lambda(\pi x + k)] \\ &< f(\pi)\gamma(\pi_0 x + k) + [1 - f(\pi)][-c - \lambda(\pi x + k)] \\ &= a(\pi)(\pi_0 x + k) - [1 - f(\pi)][c + \lambda x(\pi - \pi_0)]. \end{aligned}$$

For $\pi \geq \pi_0, PU(\pi, s; 0, 1) < a(\pi)(\pi_0 x + k) = PU(\pi, \underline{s}; p^*(\underline{s})) \leq PU^*(\pi)$. For $\pi < \pi_0$, a sufficient (but not necessary) condition for $PU(\pi, s; 0, 1) < 0 = PU^*(\pi)$ is $\pi_0 \leq c/\lambda x$.

Next consider type (2) asymmetric behavior; g randomizes using $p_g(s)$ and i rejects $s \in (0, \underline{s})$. Then

$$\begin{aligned} PU(\pi, s; p_g(s), 1) &= f(\pi)\{p_g(s)[-c + \gamma(\pi x + k)] + [1 - p_g(s)]\gamma s\} \\ &\quad + [1 - f(\pi)][-c - \lambda(\pi x + k)] \\ &= p_g(s)[a(\pi)(\pi x + k) - c] \\ &\quad + [1 - p_g(s)]PU(\pi, s; 0, 1), \end{aligned}$$

where $PU(\pi, s; 0, 1)$ is given above. It was argued above that for all $\pi, PU^*(\pi) > a(\pi)(\pi x + k) - c$ and $PU^*(\pi) > PU(\pi, s; 0, 1)$. Thus $PU^*(\pi) > PU(\pi, s; p_g(s), 1)$ for all π.

Finally, consider type (3) asymmetric behavior: g accepts s and i randomizes using $p_i(s)$, for $s \in (0, \underline{s})$. Then

$$\begin{aligned} PU(\pi, s; 0, p_i(s)) &= f(\pi)\gamma s + [1 - f(\pi)] \\ &\quad \times \{p_i(s)[-c - \lambda(\pi x + k)] - [1 - p_i(s)]\lambda s\} \\ &= [1 - p_i(s)]a(\pi)s + p_i(s)PU(\pi, s; 0, 1). \end{aligned}$$

Previous arguments imply that $PU^*(\pi) \geq a(\pi)s$ and $PU^*(\pi) > PU(\pi, s; 0, 1)$ for all π. Thus $PU^*(\pi) \geq PU(\pi, s; 0, p_i(s))$. The parametric restriction $\pi_0 \leq c/\lambda x$ was used only to show that $PU^*(\pi) > PU(\pi, s; 0, 1)$ for $\pi < \pi_0$; any other restriction which is sufficient to ensure this could be substituted.

Proof of Proposition 2. (a) Recall that $p^*(s) = 1 - \exp\{[A(\underline{s}) - A(s)]/c\}$. Note that $A(\cdot)$ also depends upon the parameters k, x, γ, and λ: $A(s; k, x, \gamma, \lambda) = \int a((s - k)/x)ds$, where $a(\cdot) = f(\cdot)\gamma - (1 - f(\cdot))\lambda$. It follows immediately that $p^{*\prime}(s) > 0$ and $\partial p^*(s)/\partial c < 0$ for $s \in (\underline{s}, \bar{s}]$. Differentiation of $p^*(s)$ with respect to any other parameter m yields the following formula:

$$\begin{aligned} \partial p^*(s)/\partial m &= -(1/c)\exp\{[A(\underline{s}) - A(s)]/c\} \\ &\quad \times [A'(\underline{s})(d\underline{s}/dm) + \partial A(\underline{s})/\partial m - \partial A(s)/\partial m]. \end{aligned}$$

Since

$$A'(\underline{s}) = a(\pi_0) = 0,$$

$$\begin{aligned} \operatorname{sgn} \partial p^*(s)/\partial m &= \operatorname{sgn}[\partial A(s)/\partial m - \partial A(\underline{s})/\partial m] \\ &= \operatorname{sgn} \partial^2 A(s)/\partial m \partial s \\ &= \operatorname{sgn} \partial a((s - k)/x)/\partial m. \end{aligned}$$

The assertions of Proposition 2 (a) then follow from the facts that

$$\partial a((s - k)/x)/\partial k = a'((s - k)/x)(-1/x) < 0,$$

$$\partial a((s - k)/x)/\partial x = a'((s - k)/x)(-1/x^2)(s - k) < 0,$$

$$\partial a((s - k)/x)/\partial \gamma = f((s - k)/x) > 0 \text{ and}$$

$$\partial a((s - k)/x)/\partial \lambda = -(1 - f((s - k)/x)) < 0.$$

(b) Recall that for $\pi \in [\pi_0, 1], s^* = \pi x + k$; the assertions follow immediately.

(c) For $\hat{p}(\pi) = 1 - \exp\{[A(\underline{s}) - A(s^*(\pi))]/c\}$, the assertions $\hat{p}'(\pi) > 0$ and $\partial \hat{p}(\pi)/\partial c < 0$ are immediate for $\pi \in (\pi_0, 1]$. Differentiating with respect to any other parameter m yields

$$\begin{aligned}\partial\hat{p}(\pi)/\partial m&\\ &= -(1/c)\exp\{[A(\underline{s})-A(s^*(\pi))]/c\}\\ &\quad\times\{A'(\underline{s})(d\underline{s}/dm)-A'(s^*)(ds^*/dm)\\ &\quad\quad+\partial A(\underline{s})/\partial m-\partial A(s^*)/\partial m\}.\end{aligned}$$

Again, $A'(\underline{s})=a(\pi_0)=0$, and $ds^*/dm=0$ for $m=\gamma$ and $m=\lambda$, so $\operatorname{sgn}\partial\hat{p}(\pi)/\partial m=\operatorname{sgn}\partial^2A(s)/\partial m\partial s=\operatorname{sgn}\partial a((s-k)/x)/\partial m$ for $m=\gamma$ and $m=\lambda$. The claims that $\partial\hat{p}(\pi)/\partial\gamma>0$ and $\partial\hat{p}(\pi)/\partial\lambda<0$ follow from the facts that $\partial a((s-k)/x)/\partial\gamma=f((s-k)/x)>0$ and $\partial a((s-k)/x)/\partial\lambda=-(1-f((s-k)/x))<0$, respectively. For $m=k,x$ there are two conflicting effects; $\operatorname{sgn}A'(s^*)(\partial s^*/\partial m)>0$, while $\operatorname{sgn}\partial a((s-k)/x)/\partial m<0$. Thus the signs of $\partial\hat{p}(\pi)/\partial k$ and $\partial\hat{p}(\pi)/\partial x$ are indeterminate at this level of generality.

Proof of Proposition 3. Substituting for s^0 in equation (5), it is clear that $U_1\geq U_2$; it is never optimal to take all cases to trial. Comparing U_1 and U_3 gives the following four cases:

Case 1: For $q\geq\max\{\lambda/(\gamma+\lambda),c/(\gamma x(E_g-E_i)+c)\}$, $s=s^0$ is optimal; the payoff is given by U_1.

Case 2: For $q\in[\lambda/(\gamma+\lambda),c/(\gamma x(E_g-E_i)+c)]$, $s=s_0$ is optimal; the payoff is given by U_3.

Case 3: For $q\in[(c+\lambda[E_ix+k])/(\gamma[E_gx+k]+\lambda[E_ix+k]+c),\lambda/(\gamma+\lambda)]$, $s=s^0$ is optimal; the payoff is given by U_1.

Case 4: For $q\leq\min\{(c+\lambda(E_ix+k])/(\gamma[E_gx+k]+\lambda[E_ix+k]+c),\lambda/(\gamma+\lambda)\}$, $s=0$ is optimal; the payoff is given by U_3.

Actually, when the interval in Case 2 is nonempty, the interval in Case 3 is empty and vice versa. This allows the simplification in Proposition 3.

Proof of Proposition 4. Define the function $\Delta(q)=\mathrm{EPU}^*(q)-\mathrm{EPU}^r(q)$, where $\mathrm{EPU}^*(q)$ is given by equation (10) and $\mathrm{EPU}^r(q)$ is given by equations (8) and (9) in cases (A) and (B), respectively, of Proposition 3. Restricted discretion is preferred if and only if $\Delta(q)<0$. We need to show that there exists a unique $q_0\in(0,1)$ such that $\Delta(q)>0$ for $q<q_0$ and $\Delta(q)<0$ for $q>q_0$. Note that $\mathrm{EPU}^*(q)$ is a linear function of q, while $\mathrm{EPU}^r(q)$ is a continuous and piecewise linear function of q. Thus $\Delta(q)$ is also a continuous and piecewise linear function of q.

Consider first case (B) of Proposition 3. $\Delta(q)=\mathrm{EPU}^*(q)>0$ for all $q\in I_{1B}$. For $q\in I_{2B}$, $\Delta(q)$ can be written

$$\text{(A11)}\quad \Delta(q)=\int_0^1[\mathrm{PU}^*(\pi)-\gamma(\pi x+k)]qd\Phi(\pi|g)$$
$$+\int_0^1[PU^*(\pi)+c+\lambda(\pi x+k)](1-q)d\Phi(\pi|i).$$

Since $\mathrm{PU}^*(\pi)=a(\pi)(\pi x+k)-\hat{p}(\pi)c<\gamma(\pi x+k)$ for $\pi>\pi_0$ and $\mathrm{PU}^*(\pi)=0<\gamma(\pi x+k)$ for $\pi\leq\pi_0$, the first integrand is negative. Since $\mathrm{PU}^*(\pi)\geq0$ for all π, the second integrand is positive. Thus $\Delta'(q)<0$ for $q\in I_{2B}$ and $\Delta(1)<0$. Recall that $\Delta(q)>0$ for $q\in I_{1B}$ and $\Delta(q)$ is continuous in q. Consequently, there exists a unique $q_0\in I_{2B}$ such that $\Delta(q)>0$ for $q<q_0$ and $\Delta(q)<0$ for $q>q_0$. This proves the claim for Case (B).

Next consider Case (A) of Proposition 3. For $q\in I_{1A}$, $\Delta(q)=\mathrm{EPU}^*(q)>0$. For $q\in I_{2A}$, $\Delta(q)$ can be written

$$\Delta(q)=\int_0^1\mathrm{PU}^*(\pi)qd\Phi(\pi|g)$$
$$-\int_0^1\gamma(\pi x+k)qd\Phi(\pi|i)$$
$$+\int_0^1[\mathrm{PU}^*(\pi)+\lambda(\pi x+k)](1-q)d\Phi(\pi|i).$$

For $q\in I_{3A}$, $\Delta(q)$ is as above in equation (A11). For $q\in I_{2A}$, $\Delta(q)$ is a linear function with $\Delta(\lambda/(\lambda+\gamma))>0$. Thus $\Delta(q)$ either remains positive throughout I_{2A} or there exists a unique $q_0\in I_{2A}$ such that $\Delta(q)>0$ for $q<q_0$ and $\Delta(q)<0$ for $q>q_0$. If $\Delta(q)$ remains positive throughout I_2, then since $\Delta(q)$ is continuous with $\Delta(1)<0$ and $\Delta'(q)<0$ for $q\in I_{3A}$, there exists a unique $q_0\in I_{3A}$ such that $\Delta(q)>0$ for $q<q_0$ and $\Delta(q)<0$ for $q>q_0$. If $\Delta(q)=0$ for some $q_0\in I_{2A}$, then since $\Delta'(q)<0$ for $q\in I_{3A}$, $\Delta(q)$ remains negative thereafter. In either case, the claim follows for Case (A).

REFERENCES

Alschuler, Albert W., "The Changing Plea Bargaining Debate," *California Law Review*, 1981, *69*, 652–730.

______, "The Prosecutor's Role in Plea Bargaining," *University of Chicago Law Review*, Autumn 1968, *36*, 50–112.

Banks, Jeffrey S. and Sobel, Joel, "Equilibrium Selection in Signaling Games," *Econometrica*, May 1987, *55*, 647–62.

Bebchuk, Lucien Arye, "Litigation and Settlement Under Imperfect Information," *Rand Journal of Economics*, Autumn 1984, *15*, 404–15.

Cho, In-Koo and Kreps, David M., "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, May 1987, *102*, 179–222.

Church, Thomas W., Jr., "In Defense of 'Bargain Justice,'" *Law and Society*, Winter 1979, *13*, 509–25.

Easterbrook, Frank H., "Criminal Procedure as a Market System," *Journal of Legal Stud-*

ies, June 1983, *12*, 289–332.

Forst, Brian and Brosi, Kathleen B., "A Theoretical and Empirical Analysis of the Prosecutor," *Journal of Legal Studies*, January 1977, *6*, 177–91.

Gould, John P., "The Economics of Legal Conflicts," *Journal of Legal Studies*, June 1973, *2*, 279–300.

Grossman, Gene M. and Katz, Michael L., "Plea Bargaining and Social Welfare," *American Economic Review*, September 1983, *73*, 749–57.

Jackson, Bruce, *Law and Disorder*, Chicago: University of Illinois Press, 1984.

Kreps, David M. and Wilson, Robert, "Sequential Equilibria," *Econometrica*, July 1982, *50*, 863–94.

Landes, William M., "An Economic Analysis of the Courts," *Journal of Law and Economics*, April 1971, *14*, 61–107.

P'ng, Ivan P. L., "Strategic Behavior in Suit, Settlement, and Trial," *Bell Journal of Economics*, Autumn 1983, *14*, 539–50.

Reinganum, Jennifer F. and Wilde, Louis L., "Settlement, Litigation and the Allocation of Litigation Costs," *Rand Journal of Economics*, Winter 1986, *17*, 557–66.

Rhodes, William M., "The Economics of Criminal Courts: A Theoretical and Empirical Investigation," *Journal of Legal Studies*, June 1976, *5*, 311–40.

Salant, Stephen W., "Litigation of Settlement Demands Questioned by Bayesian Defendants," Social Science Working Paper No. 516, California Institute of Technology, March 1984.

Samuelson, William, "Negotiation versus Litigation," Boston University School of Management Working Paper No. 7/84, March 1983.

Shavell, Steven, "Suit, Settlement and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs," *Journal of Legal Studies*, January 1982, *11*, 55–82.

Silberman, Charles E., *Criminal Violence, Criminal Justice*, New York: Vintage Books, 1980.

Spulber, Daniel F., "Negligence, Contributory Negligence and Pre-Trial Settlement Negotiation," MRG Working Paper No. M8511, University of Southern California, July 1985.

Vera Institute of Justice, *Felony Arrests: Their Prosecution and Disposition in New York City's Courts*, New York: Longman, 1981.

Weimer, David L., "Plea Bargaining and the Decision to go to Trial: The Application of a Rational Choice Model," *Policy Sciences*, 1978, *10*, 1–24.

Domestic Politics, Foreign Interests, and International Trade Policy

By Arye L. Hillman and Heinrich W. Ursprung*

This paper incorporates the foreign interest in the determination of a country's international trade policy into a model of political competition between candidates contesting elective office. We envisage foreign and domestic producer interests as expressing political support for a candidate via campaign contributions, and candidates as making trade policy pronouncements to maximize political support from producer interests. Tariffs are divisive, but VERs are consistent with conciliatory policy positions yielding mutual gain to foreign and domestic interests. No candidate has an interest in formulating a trade policy position using a tariff if a VER is a policy option.

Protection in the developed countries has in recent years often taken the form of bilaterally negotiated voluntary export restraints (VERs) rather than the more traditional unilaterally imposed import restrictions.[1] The use of voluntary export restraints has been paralleled by interest in the theoretical literature. The effects of VERs have been compared with tariffs and quotas, and VERs have been examined in a political-economy framework with a view to identifying the gainers and losers from this form of restriction of international trade.[2] Foreign interests figure prominently in these analyses. VERs are, like import quotas which have been assigned to domestic interests, quantitative restrictions, but it is foreigners who secure the rents created by the restraint on trade. If a VER is a substitute for a tariff, revenue which would have accrued to the home government is transferred to foreigners.

The rent or revenue transfer to foreigners has given rise to two interrelated perspectives on VERs. By negotiating VERs, home governments have been seen as providing compensation in seeking foreign acquiescence with domestic protectionist policies which violate prior negotiated GATT commitments; and by consenting to VERs, foreign governments have on their part been seen as preempting other more costly protectionist measures in export markets.[3]

---

*Bar-Ilan University, Ramat Gan, 52100 Israel and University of Konstanz, Konstanz, D7750 F.R.G., respectively. Ursprung acknowledges the support of the Swiss National Science Foundation. We thank Peter Bernholz, Harold Demsetz, Sebastian Edwards, Wolfgang Mayer, Michael Waldman, and two anonymous referees for comments on a previous draft. We have also benefited from seminars at Bar-Ilan University, the University of California-Los Angeles, the University of Konstanz, the University of Pennsylvania, the University of Pittsburgh, and Princeton University.

[1] The various instances include the restraints on U.S. imports of Japanese automobiles, initially implemented in 1981 and renewed in subsequent years to date. Japanese automobile exports to the U.K., France, and Italy are also subject to voluntary restraints, as are European exports of steel to the United States, various developing countries' exports of clothing to the United States, and the EEC, and exports of consumer electronics goods from the newly industrializing countries of Asia to some developed countries. VERs as a component of U.S. trade policy date back at least to 1957 when the United States announced that Japan would "voluntarily" restrict exports of cotton textiles and apparel for the subsequent five years.

[2] See Kent Jones (1984), who also reviews prior theoretical literature.

[3] Other proposed explanations for VERs include the suggestion by Jagdish Bhagwati (1985) that export restraints may be linked to incentives for domestic investment by the foreign industry whose exports are subject to restriction. Since it is Japanese exports which have been prominently involved, it has also been suggested that VERs may reflect a traditional Japanese preference for undertaking unavoidable, unpleasant acts oneself rather than leaving the execution of such acts to others.

VERs as protectionist options thus introduce a focus on the foreign interest in the domestic determination of international trade policy, and more particularly in the *form* which protectionist policies take. This paper presents a model of the political formulation of protectionist policies which takes explicit account of this foreign interest. In a setting of representative democracy, rival candidates for political office are viewed as making trade policy pronouncements so as to maximize political support, which may derive from foreign or domestic producer interests.[4]

Foreigners, lacking the necessary franchise, cannot influence the determination of trade policy directly by voting in domestic political contests. However, under representative democracy, foreign participation in domestic politics can take the form of campaign contributions, or other transfers directed at influencing the trade-policy position taken by a political candidate. Evidence confirming the foreign interest in participation in domestic politics is provided by the comprehensive account of foreign lobbying expenditures documented by Steven Husted (1986).

We shall portray the foreign interests seeking to influence the domestic formulation of international trade policy as residual claimants in a foreign export industry. The foreign industry is composed of identical Cournot firms acting noncooperatively, unless directed to comply with negotiated VERs by collectively restraining exports and preallocating market shares. A Cournot portrayal of firms is likewise adopted in the domestic import-competing industry. Firms in the domestic industry remain however at all times in compliance with domestic antitrust laws and do not behave collusively.[5]

Firms choose campaign contributions to a favored candidate to maximize expected profits, which depend upon the rival candidates' prospects for election and upon the trade policy to which a candidate commits himself if elected. The campaign contributions made by firms have a public-good character. Any one firm's outlays benefit other firms which stand to gain or lose in the same way from a candidate's policy commitment. Firms' political outlays thus constitute an instance of private provision of public goods, here the collective benefit derived from influence over policy.[6]

In this endogenous policy setting, we investigate the characteristics of protectionist proposals when the alternative forms of trade intervention are tariffs and quantitative restrictions in the form of VERs.[7] The outcome of political competition depends upon the instrument used by the rival candidates to make trade-policy pronouncements. The tariff equilibrium is politically divisive in that a protectionist candidate announces a prohibitive tariff and a liberal trade-policy candidate announces free trade as his policy platform. Associated with the political divisiveness in the candidates' policy pronouncements is divisiveness in the attribution of economic gains and losses associated with candidates' pronounced policies. The policy platform chosen by the protectionist candidate maximizes domestic profits and minimizes foreign profits, while the opposite is true of the equilibrium policy platform chosen by the liberal trade-policy candidate.

---

[4] Economic policy is thus politically determined (see, for example, Peter Bernholz, 1974, 1977). The international economics literature has in recent years emphasized the inherently political nature of protectionist decisions. See, for example, Robert Baldwin, 1982, 1984, 1985; Arye Hillman, 1982; Wolfgang Mayer, 1984; Leslie Young and Stephen Magee, 1986; and James Cassing and Hillman, 1985, 1986. Our model of political competition under representative democracy is in particular of the genre of models of political contestability as formulated by Young and Magee. The literature on the political economy of trade policy is reviewed in Hillman, 1988.

[5] Firms are viewed as identical, so that the number of firms can be used as a simple measure of industry competitiveness. More generally, of course, because of efficiency differences (see Harold Demsetz, 1973), firms will have different market shares. However, we abstract from such differences among firms to focus on the quest to influence trade policy via the political process.

[6] See Richard Cornes and Todd Sandler (1986) for an exposition of the theory of noncooperative voluntary provision of public goods.

[7] We do not compare auctioned import quotas with VERs because under our assumptions (which we shall subsequently make explicit) concerning property rights to revenue assignment, political candidates prefer VERs to auctioned quotas as protectionist instruments. The literature offers other perspectives on political choice of the means of protection. The choice between tariffs and quotas by a government trading off political support from domestic gainers and losers from protection is considered by Cassing and Hillman, 1985. On terms of trade uncertainty, domestic interests, and the political choice between tariffs and quotas, see Rodney Falvey and Peter Lloyd, 1986. Again in settings concerned only with the domestic interests influenced by trade policy, Dani Rodrik, 1986, and Mayer and Raymond Riezman, 1987, consider aspects of the choice between tax/subsidy instruments and tariffs.

However, when candidates' policy proposals are formulated in terms of export restraints, trade-policy platforms converge to a Hotelling-type equilibrium. There is thus political conciliation. At the same time, the prospect emerges of compromise between the conflicting economic interests of domestic and foreign producers. An interior equilibrium (which cannot arise under a tariff) is characterized by a level of export restraint which is more restrictive than foreigners would wish in order to maximize collusively joint profits, but less restrictive than the prohibitive level of restraint of imports sought by domestic import-competing firms.[8]

Beyond compromise, the political equilibrium associated with VERs is consistent with mutual economic gain to foreign and domestic producer interests.[9] Our conception of voluntariness of export restraints is based on the prospect of such mutual gain. We shall term an export restraint as voluntary if foreign and domestic industry rents both increase as the consequence of foreigners' coordinated restriction of export sales. This conception of voluntariness differs from the more traditional perspective that export restraints are imposed restrictions which are only voluntary in the sense that foreigners find negotiation and compliance preferable to a threatened alternative trade restriction.[10]

Whether an export restraint facilitates increases in producer rents at home and abroad will be shown to be determined by a relationship linking foreign and domestic market structures and the substitutability of foreign and domestic goods in domestic consumption. In particular, a change in substitutability as for example effected by upgrading of exports can make an initially involuntary constraint voluntary.[11]

We also investigate the political choice between VERs and tariffs. Export restraints are shown to dominate tariffs, in that political equilibria are inconsistent with any candidate's making a protectionist policy pronouncement in terms of a tariff when the option of using export restraints is available.

Two aspects of our model of the political contestability of trade policy merit preliminary comment. We assume that rival candidates place no value on revenue derived from a tariff. Such revenue accrues to the general fund of government receipts to which the candidates have no claim.[12] However, while the candidates have no means of appropriating or benefiting from the revenue from a tariff which they might propose, there is a prospective benefit from the creation of rents which can be transferred to private economic interests via export restraints. The latter rents can be transformed via the political process into campaign contributions, thereby bearing upon the candidates' ultimate concern, their probabilities of attaining political office.

---

[8] For consistent evidence, see Schlesinger, *Wall Street Journal*, Jan. 28, 1987, "Japan's Extension of Auto-Export Curb Fails to Mollify Strongest U.S. Critics."

[9] For recognition of collusive aspects of VERs, see Richard Harris, 1985; Jurgen Eichenberger and Ian Harper, 1987, and Kala Krishna (forthcoming 1988).

[10] There is a related view in the literature (see Bhagwati and T. N. Srinivasan, 1976) that foreigners may themselves initiate voluntary restraints on exports to preempt protectionist measures in export markets. Consistent with this view, the *Economist* (May 30, 1987) reported: "Taiwan and South Korea may be about to introduce voluntary export restraints on some goods, including shoes. Legislation in the United States Congress is trying to limit shoe imports to their 1986 levels."

[11] Thus the United States indicated its readiness in 1985 to dispense with the voluntary export restraints on Japanese automobiles which had been introduced in 1981. However, the voluntary export restraints were nonetheless renewed by the Japanese government. At the same time there is evidence of quality upgrading of Japanese automobile exports to the United States (see Robert Feenstra, 1984, 1985; and Charles Collyns and Steven Dunaway, 1987).

[12] See Baldwin (1984) for observations on limitations to the assignment of property rights to tariff revenues. For an analysis encompassing voluntary export restrictions which does on the other hand assume that individuals can secure claims to revenue from tariffs, see Richard Brecher and Bhagwati (1987).

Second, our formulation of political competition does not assign an active role to consumer interests. The focus is on those interests with the principal economic stakes in influencing the political formulation of trade policy, domestic and foreign residual claimants who benefit or lose according to the trade policy adopted. Consumers are viewed as passively influenced in their voting behavior by campaign expenditures, hence the importance assigned to campaign contributions by the candidates. The vote which consumers cast is, moreover, preconditioned on the policy positions taken by the candidates on a range of issues and is not dominated by any one single issue. Out of "rational ignorance," consumer–voters have no reason to have acquired detailed information on the consequences of policies proposed regarding the extent and form of protection for a particular industry. This is, however, not to deny that consumer interests can influence the trade policies proposed by candidates for political office. To the extent that consumer interests matter, the protectionist outcomes which we shall portray will be moderated.

I. The Model

We consider an import-competing industry producing a homogeneous good which is imperfectly substitutable in domestic consumption for imports. The domestic demand functions for the domestic good and imports are

$$P = a - bx + \gamma P^* \tag{1}$$

$$P^* = a - bx^* + \gamma P, \tag{2}$$

where x and x^* denote quantities of domestically produced and foreign goods, and P and P^* are the respective domestic prices of the two goods; γ is a constant with values $0 \le \gamma < 1$ which measures substitutability in consumption between domestic and foreign goods.[13]

The domestic industry consists of n identical profit-maximizing Cournot firms. Each firm makes a campaign contribution L_i to influence the outcome of an electoral contest between two candidates, one of whom is predisposed to a protectionist and the other to a liberal trade policy.[14] Firms' per unit production costs are a constant c. A domestic firm chooses output and a campaign contribution to the protectionist candidate to maximize profits given by

$$\max_{x_i,\, L_e} \pi_i = (P - c)x_i - L_i, \qquad i = 1, \ldots, n. \tag{3}$$

The foreign industry consists of m Cournot firms also seeking to influence domestic trade policy. Foreign firms have the same per unit cost of production as domestic firms. When a tariff is the means of domestic protection, the profit-maximization problem confronting a foreign firm is

$$\max_{x_i^*,\, L_i^*} \pi_i^* = (P^* - c - t)x_i^* - L_i^*, \qquad i = 1, \ldots, m, \tag{4}$$

where t is a specific tariff. Alternatively, with an export restraint V, foreign firms maximize profits subject to the coordinated market-sharing condition

---

[13] If $\gamma = 0$, demand for each good depends only upon its own price, reflecting zero substitutability. With $\gamma > 0$, if $P = P^*$, then $x = x^*$; but $P > P^*$ is consistent with $x > 0$, so not all domestic firms' sales are lost to competitive imports when the price of the domestic good exceeds that of imports. For the specification of consumption preferences underlying the above demand functions and another application to an analysis of export restraints, see Harris (1985). Using Bertrand price adjustment (as distinct from our formulation of Cournot quantity-adjustment), Harris shows that an export restraint set at the free-trade quantity of imports can be voluntary, in the sense of facilitating an increase in foreign firms' profits. See also Eichenberger and Harper, 1987, and Krishna, forthcoming 1988.

[14] Ideology has no role here. We assume a policy predisposition to distinguish the candidates.

$$(5)\qquad x_i^* \le V/m.$$

Each candidate seeks to maximize his probability of election, which is determined by campaign contributions received relative to total contributions. The protectionist candidate, who is the beneficiary of domestic firms' contributions, thus seeks to maximize

$$(6)\qquad W = \frac{L}{L + L^*},$$

where

$$L = \sum_{i=1}^{n} L_i \quad \text{and} \quad L^* = \sum_{i=1}^{m} L_i^*.$$

On the other hand, the candidate predisposed to a liberal trade policy seeks to minimize W. It will be assumed that if the two candidates propose identical policies, then each has equal probability of election.

II. Political Competition with a Tariff

A. *Tariffs and Profits*

Let candidates' policy pronouncements be made in terms of tariffs. For given tariff policies, each firm maximizes profits, assuming that other firms do not respond to its output choice. The symmetric Nash equilibrium that results is given by the respective outputs for domestic and foreign firms,

$$(7)\quad x_i = \frac{[m(1-\gamma)+1]B}{bA} + \frac{[\gamma(1-\gamma^2)m]t}{bA}, \quad i = 1,\dots,n$$

and

$$(8)\quad x_i^* = \frac{[n(1-\gamma)+1]B}{bA} - \frac{(1-\gamma^2)(n+1)t}{bA}, \quad i = 1,\dots,m$$

where $A \equiv (m+1)(n+1) - mn\gamma^2 > 0$ and $B \equiv (1+\gamma)(a - c(1-\gamma)) > 0$.

Denote by $\bar{t}$ the prohibitive tariff in a symmetric Nash equilibrium, established by solving (8) for $x_i^* = 0$. For $x_i, x_i^* \ge 0$, $\gamma > 0$ and $0 \le t \le \bar{t}$, (7) and (8) indicate that equilibrium domestic sales are linear functions of the tariff, increasing for domestic firms and decreasing for foreign firms. As the tariff is increased, domestic consumers substitute domestic output for imports, the extent of substitution depending on the value of γ. If $\gamma = 0$, the output x_i of domestic firms is independent of the tariff, but the domestic sales x_i^* of foreign firms still decline as the tariff is increased.

Post-election profits depend upon the tariff via[15]

$$(9)\quad \pi_i(t) = \frac{1}{b(1-\gamma^2)A^2} \times \{[m(1-\gamma)+1]B + (1-\gamma^2)\gamma m t\}^2$$

$$(10)\quad \pi_i^*(t) = \frac{1}{b(1-\gamma^2)A^2} \times \{[n(1-\gamma)+1]B - (n+1)(1-\gamma^2)t\}^2.$$

It follows that as depicted in Figure 1, for nonprohibitive tariffs, profits of domestic and foreign firms are respectively strictly increasing and decreasing, but both strictly convex, functions of the tariff.

B. *Choice of Campaign Contributions*

Perceiving the *ex post* (or post election) dependence of profits on the tariff, firms make campaign contributions *ex ante* to influence political candidates' trade policy pro-

---

[15] To derive the profit functions (9) and (10), substitute the equilibrium quantities (7) and (8) into (3) and (4), respectively.

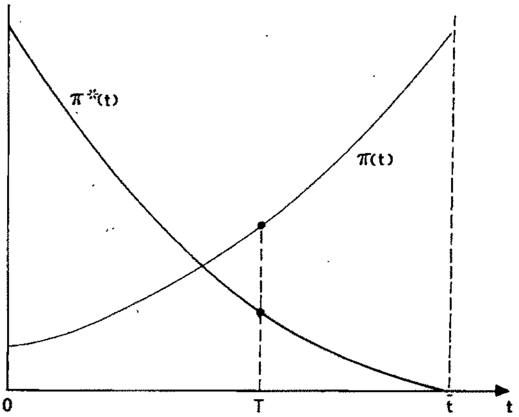


FIGURE 1

nouncements. Firms choose their individual contributions as Nash equilibria. The outlays made reflect the public-good nature of the benefits to other firms in the industry. The value of campaign contributions is a deduction from *ex post* profits given by (9) and (10).[16]

When deciding on campaign contributions, firms confront the uncertainty of not knowing which candidate will be elected. Firms are risk-neutral and choose campaign contributions to maximize expected profits. For a domestic firm, expected profits are

$$(11)\quad E\pi_i = \theta[\pi_i(t_0) - L_i] + (1-\theta)[\pi_i(t_1) - L_i] = \theta\pi_i(t_0) + (1-\theta)\pi_i(t_1) - L_i,$$

where θ and $(1-\theta)$ are the respective probabilities of election of the liberal trade-policy and protectionist candidates, and t_0 and t_1 $(t_0 \le t_1)$ are the candidates' respective tariff pronouncements. The expected profits of a foreign firm, $E\pi_i^*$, likewise encompass the two states of the world described by the policy pronouncements of the rival candidates, and take the same form as (11).

Campaign contributions determine the candidates' probabilities of electoral success, via

$$(12)\quad \theta = L^*/(L+L^*) = W(L, L^*)$$
$$1-\theta = L/(L+L^*) = 1 - W(L, L^*),$$

where L and L^* are as defined in (6) total contributions by domestic and foreign firms.

On substituting (12) into (11), it follows that domestic firms choose campaign contributions which satisfy

$$(13)\quad \frac{\partial E\pi_i}{\partial L_i} = \frac{L^*}{(L+L^*)^2} \times [\pi_i(t_1) - \pi_i(t_0)] - 1 = 0, \quad i = 1,\dots,n.$$

The condition (13) can be expressed as

$$(14)\quad \frac{L^*}{(L+L^*)^2}\Delta\pi_i = 1, \quad i = 1,\dots,n$$

where $\Delta\pi_i = [\pi_i(t_1) - \pi_i(t_0)]$ is the difference in domestic firms' profits for the states associated with the rival candidates' tariff pronouncements.

For foreign firms, the expected profit-maximizing choice of campaign contributions satisfies a similar condition

$$(15)\quad \frac{L}{(L+L^*)^2}\Delta\pi_i^* = 1, \quad i = 1,\dots,m.$$

Combining (14) and (15) yields the relationship between equilibrium campaign contributions

$$(16)\quad \frac{L}{L^*} = \frac{\Delta\pi_i}{\Delta\pi_i^*} = \frac{\pi_i(t_1) - \pi_i(t_0)}{\pi_i^*(t_0) - \pi_i^*(t_1)} \equiv R(t_0, t_1).$$

---

[16] Observe the relation here to the portrayals of contestability of rents in the literature (for example, Hillman and Eliakim Katz, 1984; and Hillman and John Riley, 1987). The outlays L_i made in seeking to influence the outcome of the political contest (and thereby the political allocation of producer rents) are irretrievably lost.

C. *Candidates' Policy Pronouncements*

The rival candidates choose policy positions to maximize their respective probabilities of electoral success, as expressed by (12).[17] The protectionist candidate accordingly seeks to maximize $L/(L+L^*)$. This is however equivalent to maximizing L/L^*, and thus entails adopting a policy position which maximizes $R(t_0, t_1)$ as defined in (16). Hence, maximization of the protectionist candidate's probability of electoral success implies choice of a tariff which maximizes $\Delta\pi_i/\Delta\pi_i^*$, the state-contingent difference in domestic relative to foreign profits.

The relation (16) therefore links candidates' political motives and firms' profit-maximizing objectives. Each candidate has a political interest in maximizing the gains which his constituency secures from his announced policy relative to the gains which his political opponent's constituency secures from the opponent's policy.[18]

D. *Equilibrium Tariff Pronouncements*

A candidate's choice of tariff pronouncement evidently affects the profits which domestic and foreign firms would secure on his election. An increase in the tariff proposed by the protectionist candidate may increase $\Delta\pi_i$, but at the same time may also increase $\Delta\pi_i^*$. One cannot therefore presume that the protectionist candidate will necessarily announce a prohibitive tariff or that his rival will announce a policy of free trade. This is however the equilibrium outcome. The result of political competition via tariff pronouncements is described by

PROPOSITION 1: *The equilibrium platform combination when political-support maximizing candidates for political office make policy pronouncements in terms of tariffs is characterized by the protectionist candidate's announcing a prohibitive tariff and the liberal trade-policy candidate's announcing a policy of free trade.*

PROOF:

Let the candidates' policy pronouncements differ. Evaluating $\Delta\pi_i$ and $\Delta\pi_i^*$ when $t_0 < t_1$ and substituting into (16) yields

$$R(t_0, t_1)\big|_{t_0<t_1} = \frac{2[m(1-\gamma)+1]m\gamma B + m^2\gamma^2(1-\gamma^2)(t_0+t_1)}{2[n(1-\gamma)+1](n+1)B-(n+1)^2(1-\gamma^2)(t_0+t_1)}. \tag{17}$$

The expression (17) is strictly increasing in both t_0 and t_1. Hence the protectionist candidate increases his probability of election by announcing increasingly greater tariffs, up to $\bar{t}$. Conversely, the liberal trade-policy candidate increases his probability of election by announcing ever smaller tariffs, until free trade.

Now suppose that the candidates were to announce the same positive nonprohibitive tariff T, where as in Figure 1 $0 < T < \bar{t}$. (17) (which is derived subject to $t_0 < t_1$) then no longer defines election probabilities. Since identical tariff pronouncements imply equal state-contingent profits ($\Delta\pi_i = \Delta\pi_i^* = 0$), R in (16) is also not defined. However, since identical policy pronouncements imply equal probabilities of election,

$$W = \frac{L}{L+L^*} = \frac{R}{1+R} = \frac{1}{2}, \tag{18}$$

and $R(T,T) = 1$. The liberal trade-policy candidate improves his prospects of electoral success by deviating from T to announce free trade if $R(O,T) < R(T,T) = 1$, which

---

[17] The candidates for political office are thus Stackelberg leaders with respect to the domestic and foreign interests making campaign contributions.

[18] Satya Das (1988) derives a condition similar to (16). Das presents an alternative scenario which recognizes the motives of foreigners in seeking to influence domestic policy decisions. The model presented by Das has elements in common with our formulation. However, rather than rival candidates engaged in political competition making political-support maximizing trade-policy pronouncements, Das views domestic and foreign firms as seeking to influence a single policymaker's decision, and whereas we compare tariffs and voluntary export restraints as alternative instruments of trade restriction, Das focuses on the policymaker's response in choosing to implement an import quota.

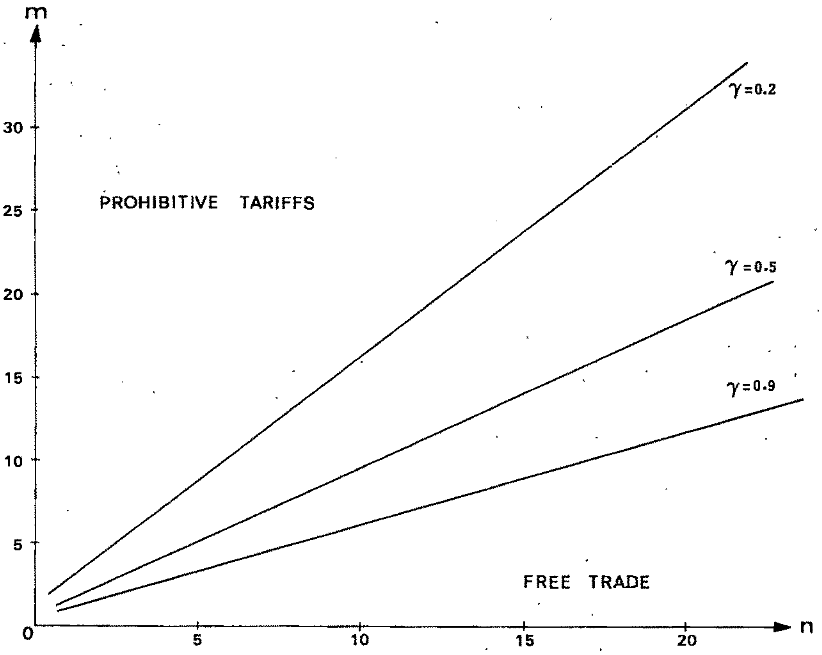


FIGURE 2

from (18) requires that

$$(19)\quad T<\frac{2B[(n+1)(n(1-\gamma)+1)-\gamma m(m(1-\gamma)+1]}{(1-\gamma^2)(m^2\gamma^2+(n+1)^2)} \equiv \bar{t}.$$

The protectionist candidate increases his probability of election by deviating from the common pronouncement T to announce a prohibitive tariff if $R(T,\bar{t})>R(T,T)=1$, which from (17) and (19) implies $T>\tilde{t}-\bar{t}\equiv\hat{t}$. If $\tilde{t}\leq\bar{t}$, then $\hat{t}\leq 0$ and this latter condition is satisfied for all $T\in(0,\bar{t})$. If however $\tilde{t}>\bar{t}$, then (19) is satisfied for all $T\in(0,\bar{t})$. Hence at least one of the candidates can improve his election prospects by deviating from the common pronouncement T, which therefore cannot be an equilibrium. We have thus eliminated a common interior Hotelling-type policy platform as an equilibrium; and we have demonstrated that when tariff pronouncements differ, the equilibrium platform consists of free-trade and prohibitive tariff pronouncements.

The tariff equilibrium is accordingly politically divisive, and is also maximally conflicting in terms of the state-contingent outcomes for foreign and domestic producer interests. A candidate's policy proposal either maximizes domestic profits by altogether excluding imports from the domestic market, or serves foreign interest by permitting free trade.[19]

E. *Comparative Statics of the Tariff Equilibrium*

Which of the two candidates will enjoy the greater likelihood of electoral success, and therefore which policy will have the greater likelihood of implementation, is determined by the exogenous characteristics of market structure and substitutability in domestic consumption between imports and domestic goods.

---

[19] In equilibrium, one candidate will in general have a higher probability of success than the other. The candidate with the lower probability of success could duplicate the policy of his opponent. Then one equilibrium policy emerges as supported by both candidates. The characteristic of no-compromise in the tariff equilibrium is retained.

Figure 2 depicts three contours (which are necessarily increasing and nonintersecting) along which $R(0,\bar{t})=1$ and hence along which free trade and protectionist outcomes are equally likely.[20] As the value of the substitution parameter γ increases, the zone wherein the protectionist candidate has the greater probability of election expands. The greater is γ, the more substitutable are domestically produced goods for imports, and hence the greater the gain to domestic interests from protection relative to free trade. Domestic interests accordingly have a greater incentive to make campaign contributions to the protectionist candidate. At the same time, foreign interests have less to lose from a denial of access to the domestic market, and make smaller campaign contributions to their candidate.

Increased competition, whether in terms of more competitors in a firm's own industry or abroad, decreases firms' profits, but *ceteris paribus* profits are decreased more by increased competitiveness in a firm's own industry, since domestic goods and imports are only imperfectly substitutable. As competition increases, profits fall, and the voluntary public-good outlays for campaign contributions fall. An increase in the number of domestic firms therefore reduces the benefits to domestic firms from a protectionist outcome by more than the reduction in benefit to foreign firms from a free-trade outcome. Hence, in Figure 2, for given γ, an increase in n holding m constant increases the probability of election of the free-trade candidate.

III. Export Restraints

Now let the candidates for political office formulate their policy pronouncements in terms of export restraints. We begin by establishing the relation between export restraints and firms' profits.

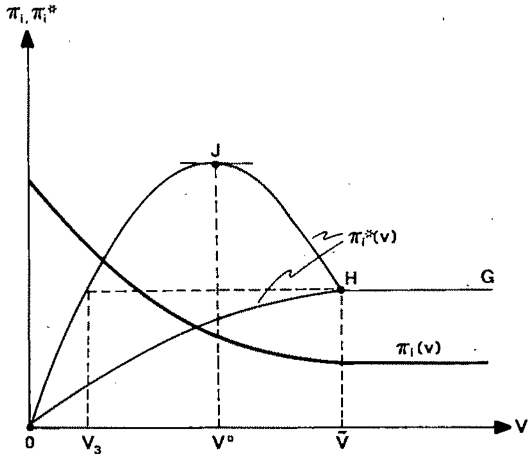


FIGURE 3

A. *Domestic Profits*

When foreign firms confront the binding export constraint V, the equilibrium output of domestic firms is

$$(20)\qquad x_i=\frac{B}{b(n+1)}-\frac{\gamma V}{n+1}$$

and profits are

$$(21)\qquad \pi_i(V)=\frac{[B-b\gamma V]^2}{b(n+1)^2(1-\gamma^2)}.$$

Let $\tilde{V}$ denote the free-trade quantity of imports. For export restraints less than $\tilde{V}$, $\pi_i(V)$ is a strictly decreasing convex function. Hence, as depicted in Figure 3, $\max \pi_i(V)=\pi_i(0)$, confirming that domestic profits are maximized by an export restraint which excludes competitive imports from the domestic market.

The profits of a domestic firm in (21) depend upon domestic market structure (via n) but are independent of foreign market structure (i.e., m). Export restraints thus eliminate the dependence, present under a

[20] $R(0,\bar{t})=[2\gamma m(m(1-\gamma)+1)]/[(n+1)(n(1-\gamma)+1]+(\gamma^2 m^2)/(n+1)^2$ which follows from (17) after appropriate substitution.

tariff, of domestic profits on foreign market structure.

B. *Foreign Profits*

A binding export constraint leads foreign firms to decrease their domestic sales from the level of free trade to the constrained quantity $x_i^* = V/m$. Profits subject to the export restraint are

$$(22) \quad \pi_i^*(V) = \frac{V[(n(1-\gamma)+1)B - b(n(1-\gamma^2)+V]}{m(n+1)(1-\gamma^2)},$$

where

$$(23) \quad \frac{\partial \pi_i^*}{\partial V} = \frac{\{B(n(1-\gamma)+1-2b[n(1-\gamma^2)+1]\}}{m(n+1)(1-\gamma^2)} \gtreqless 0.$$

Hence, while domestic profits necessarily fall, foreign profits may increase or decrease as an export restraint is marginally relaxed.

The export restraint which maximizes foreign profits and which foreign interests would therefore seek to have imposed is derived from (23) as

$$(24) \quad V^0 = \min\left\{\frac{B[n(1-\gamma)+1]}{2b[n(1-\gamma^2)+1]}, \tilde{V}\right\}.$$

There are two possibilities encompassed in (24). If foreigners can only lose from a restriction of exports, V^0 is given by free-trade domestic sales. In that case, the foreign profit function in Figure 3 is of the form *OHG*. Alternatively, joint profit maximization by foreign firms may require restricting exports below the free-trade quantity. V^0 is then given by the first expression in braces in (24) and the profit function is of the form *OJH*.

C. *Voluntariness*

We shall term an export restraint as "voluntary" if a foreigners' reduction in domestic sales increases foreign profits. From (23) voluntariness requires

$$(25) \quad m > \frac{n+1}{n(1-\gamma^2)+1}.$$

Should this condition hold, foreign and domestic firms have a common interest in restricting trade.

The right-hand side of (25) exceeds unity as long as there is substitutability in domestic consumption between domestic and foreign goods ($\gamma > 0$). (25) therefore confirms that if the foreign industry were to consist of a single firm or were collusively organized as a joint profit-maximizing cartel ($m = 1$), an export restraint could only be involuntary.

Increased competitiveness in the foreign industry can make an initially involuntary export restraint voluntary. The voluntary restraint affects the domestic market in the manner of a collusive restriction of sales: a more competitive foreign industry secures greater benefit from the implicit collusion.

The right-hand side of (25) is increasing in the substitution parameter γ. Hence, if, for example, as a consequence of upgrading imports become closer substitutes for domestic output, a voluntary restraint can become involuntary.

D. *Candidates' Export-Restraint Pronouncements*

Consider now the behavior of the rival candidates for political office. We shall demonstrate

PROPOSITION 2: *When the candidates choose levels of export restraint to maximize their respective probabilities of election, the political equilibrium is characterized by the announcement of a common policy.*

PROOF:

The proof proceeds along the same general line as that demonstrating the nature of the political equilibrium with tariffs. Let V_1 be the level of restraint announced by the protectionist candidate, and let the liberal trade-policy candidate announce V_0 (where $V_0 \geq V_1$). The candidates' political objectives and firms' profits are interconnected via

$$(26) \quad \frac{L}{L^*} = \frac{\Delta\pi_i}{\Delta\pi_i^*} \equiv R(V_0, V_1),$$

where $\Delta\pi_i$ and $\Delta\pi_i^*$ are established via (21) and (22). Let the candidates' policy pronouncements differ (with $V_0 > V_1$). Then substituting from (21) and (22) into (26) establishes that

$$(27)\quad R(V_0,V_1)|_{V_0>V_1} = \frac{m\gamma[b\gamma(V_0+V_1)-2B]}{(n+1)\left[b\left(n(1-\gamma^2)+1\right)(V_0+V_1)-(n(1-\gamma)+1)B\right]},$$

which is increasing in V_0 and V_1. The liberal trade-policy candidate therefore improves his election prospects by decreasing V_0 and the protectionist candidate improves his prospects by increasing V_1. Thus when policy pronouncements differ, V_0 and V_1 converge.

Let the policies converge to V^*. (27) is defined for $V_0 > V_1$. In particular $R(V^*,V^*)$ can therefore not be evaluated using (27). However, since, when the same policy is announced, candidates have the same probability of election, $R(V^*,V^*)=1$.

The common pronouncement V^* is an interior equilibrium if and only if, for $\varepsilon > 0$ arbitrarily small,

$$(28)\quad R(V^*+\varepsilon,V^*) \geq R(V^*,V^*)=1,$$

$$(29)\quad R(V^*,V^*-\varepsilon) \leq R(V^*,V^*)=1.$$

If V^* prohibits imports (i.e., $V^*=0$) (28) alone needs to be satisfied; if V^* entails no restriction of trade (i.e., $V^*=\tilde{V}$), then (29) alone needs to be satisfied.

$$\text{Let}\quad g(V) \equiv \lim_{\varepsilon\to 0} R(V+\varepsilon,V) = \lim_{\varepsilon\to 0} R(V,V-\varepsilon).$$

Then (28) and (29) imply

$$(30)\quad g(V^*) = \frac{2m\gamma[B-b\gamma V^*]}{(n+1)\left[n(1-\gamma)+1)B-2b\left(n(1-\gamma^2)+1\right)V^*\right]}$$

$$=1 \qquad \text{for } 0<V^*<\tilde{V}$$

$$g(V^*)\geq 1 \qquad \text{for } V^*=0$$

$$g(V^*)\leq 1 \qquad \text{for } V^*=\tilde{V}.$$

The function $g(V)$ is depicted in Figure 4.[21]

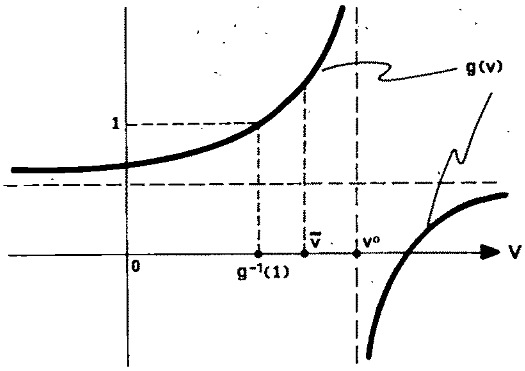


FIGURE 4

If, as depicted in Figure 4, $g(0)<1$ and $g^{-1}(1)<\tilde{V}$, there exists a unique interior solution to (30) given by

$$(31)\quad V^* = g^{-1}(1) = \frac{B[(n+1)(n(1-\gamma)+1)-2m\gamma]}{2b\left[(n+1)\left(n(1-\gamma^2)+1\right)-m\gamma^2\right]}.$$

Alternatively, if $g(0)<1$ but $g^{-1}(1)\geq\tilde{V}$, the solution to (31) is $V^*=\tilde{V}$ and the unique equilibrium is free trade. If $g(0)\geq 1$, the unique solution is prohibition of imports, or $V^*=0$.

E. *Political Compromise and Economic Gain*

When policy pronouncements are made in terms of export restraints, candidates thus announce a common policy. But to whose advantage? Clearly, if political competition results in a policy pronouncement prohibiting imports, the outcome is to the exclusive advantage of domestic import-competing interests. Or, if the political equilibrium entails free trade, and if at the same time foreigners can only lose from a curtailment of their domestic sales, the outcome is to the exclusive advantage of foreign interests. In these latter instances, the economic division of gains and losses from pronounced export-restraint policies is no different from that under tariffs. The export-restraint equilibrium is politically conciliatory in that both candidates announce the same policy, but still as with the tariff the equilibrium is divi-

---

[21] $g(V)$ converges to infinity at the regular maximum of $\pi_i^*(V)$ if $V^0<\tilde{V}$. $g(0)=2m\gamma/(n+1)[n(1-\gamma)+1] > 0$; if $g(0)<1$, then V^* is necessarily nonprohibitive.

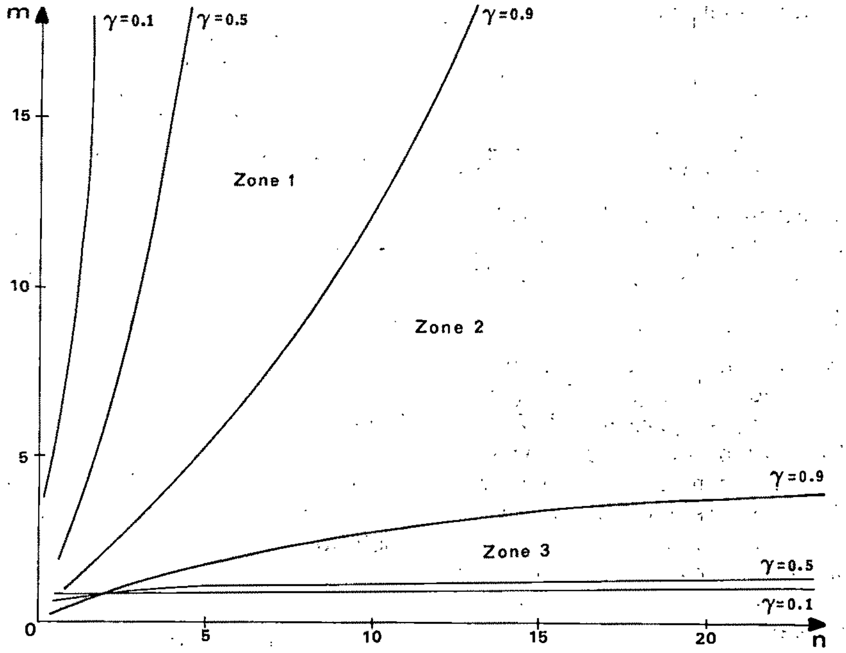


FIGURE 5

sive in offering no compromise between foreign and domestic economic interests.

The distinguishing characteristic of export-restraint equilibria is however that political competition is consistent with interior solutions. Such solutions may or may not benefit foreign interests. Whether foreign interests can gain is determined by whether condition (25) holds.

Suppose that (25) is satisfied. So export restraints can benefit foreigners. For any level of restraint announced by the protectionist candidate, let the liberal trade-policy candidate announce as his policy the foreign collusive profit-maximizing level of restraint V^0. However,

$$(\partial\pi_i^*/\partial V)|_{V^0}=0 \quad \text{and} \quad (\partial\pi_i/\partial V)|_{V^0}<0.$$

Hence this latter candidate, in seeking to minimize $R=\Delta\pi_i/\Delta\pi_i^*$, can gain by marginally decreasing his pronouncement below V^0. The liberal trade-policy candidate has thus an incentive to deviate from the foreign joint profit-maximizing level of export restraint to propose a more restrictive trade policy than sought by foreign interests. The protectionist candidate similarly has an incentive to deviate from a policy pronouncement of prohibition of imports, to announce a policy allowing some imports, which is then inconsistent with maximizing domestic firms' profits. The candidates' policy pronouncements converge in the manner we have described. The political equilibrium is a compromise between the outcomes sought by foreign and domestic interests, but also provides mutual gain to both sets of producer interests.

The compromise character of an export-restraint equilibrium is likewise reflected in an interior equilibrium where mutual gain is absent. Although foreigners lose, they lose less than they would were the equilibrium policy exclusively to accommodate domestic interests.

F. *Comparative Statics*

Comparative static properties of export-restraint equilibria are described in Figure 5. In region 1 $g(0)=2m\gamma/(n+1)(n(1-\gamma)+1)$

≥ 1 and the political equilibrium yields an export restraint which is prohibitive. In regions 2 and 3, $g(0) < 1$ and the equilibrium policy allows imports. For a given number of domestic firms, the outcome is free trade (region 3) if the foreign industry is sufficiently concentrated. In region 2 characterizing interior equilibria, the equilibrium export restraint is more stringent the more competitive is the foreign export industry, since from (31) $\partial V^*/\partial m < 0$. Also, since $\partial V^*/\partial n > 0$, the equilibrium export restraint is conversely less stringent, the more competitive is the domestic industry.

The boundaries in Figure 5 are determined by substitutability between domestic goods and imports in domestic consumption. As γ decreases, regions 1 and 3 diminish and region 2 which is consistent with interior solutions correspondingly expands. Region 1 diminishes because with decreased substitutability between imports and domestic goods a restriction on imports is of less value to domestic firms. In the limit as $\gamma \rightarrow 0$, domestic firms have no interest in restraining imports, and hence region 1 disappears. The foreign interest in influencing trade policy however remains. Indeed, as γ falls, the foreign gain from domestic market access increases. Foreigners seek the imposition of the collusive foreign profit-maximizing export restraint, which whatever the values of n and m does not entail free trade. Hence, in the limit as $\gamma \rightarrow 0$, region 3 also disappears.

IV. Political Choice of Instrument

The choice between an export restraint and a tariff as the means of protection is also a choice between assigning rents to foreigners and revenue to government. Our specification of political optimization supposes that individual candidates for political office do not themselves value the revenue derived from any tariff which they might propose. The Treasury has rights to tariff revenue. Candidates for political office can only benefit from tariff revenue via discretion in budgetary allocation—contingent on election—and then such discretion via budgetary allocations is exercised only over some small share of the revenue from any particular tariff. However, while the candidates have very limited scope for allocation of tariff revenue, their announced export-restraint policies directly influence the transfer of foreigners and the allocation of rents between domestic and foreign interests. The rent creation and rent allocation is reflected in campaign contributions received.

Because of the rent transfer, a candidate seeking support from foreign interests evidently has reason to prefer an export restraint to a tariff as the instrument for formulating his policy proposal. However, need the same be true for the protectionist candidate seeking support from domestic import-competing interests? Since it is the difference between rents secured by domestic and foreign interests in different policy regimes which determines campaign contributions and hence probabilities of election, one might propose that the protectionist candidate also prefers, given the option, to formulate his policy proposal in terms of an export restraint rather than a tariff. This is confirmed by

PROPOSITION 3: *When candidates have the choice of making policy pronouncements in terms of export restraints or tariffs, both candidates choose export restraints. The political equilibrium in the presence of a tariff option is that obtained when only export restraints are feasible.*

PROOF:

Combinations of tariffs and export restraints among which domestic firms are indifferent follow from equating profits $\pi_i(t)$ and $\pi_i(V)$ as given by (9) and (21) to obtain

$$(32) \quad V = \frac{B}{b\gamma} - \frac{n+1}{b\gamma A} \times \{[m(1-\gamma)+1]B + (1-\gamma^2)\gamma m t\}.$$

The linear function (32) is depicted as GH in Figure 6. The point G represents free trade, with either $V = \tilde{V}$ or $t = 0$; at H, there are no competitive imports and either $V = 0$ or $t = \bar{t}$. In zone 1, $\pi_i(t) > \pi_i(V)$ and domestic firms thus support a candidate announcing a tariff;

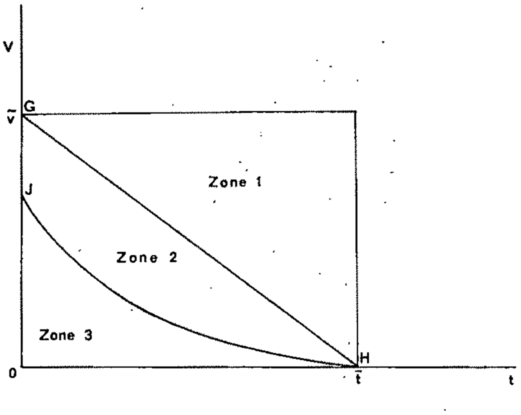


FIGURE 6

in zones 2 and 3 $\pi_i(t) < \pi_i(V)$, so domestic firms support a candidate announcing an export restraint.

JH is the corresponding policy indifference locus for foreign firms, when export restraints can increase foreign profits. (If foreign profits cannot be increased by an export restraint, *J* coincides with *G*; otherwise *J* lies below *G*, since some restraint below $\tilde{V}$ (at *G*) is beneficial for foreign firms.) To establish *JH*, use (10) and (22) to equate $\pi_i^*(t)$ and $\pi_i^*(V)$, which yields the convex relation depicted. In zones 1 and 2, $\pi_i^*(t) < \pi_i^*(V)$ and hence foreign firms support the candidate announcing an export restraint. In zone 3 $\pi_i^*(t) > \pi_i^*(V)$, so foreign firms support the candidate announcing a tariff.

In zone 2 the candidate announcing a tariff thus is supported by neither foreign nor domestic interests. In zone 1 foreign interests support the candidate proposing an export restraint and domestic interests the candidates proposing a tariff. However, the export-restraint candidate can secure the support of domestic interests by announcing a policy which moves him into zone 2. In zone 3 foreign interests support the tariff candidate and domestic interests the export-restraint candidate, but similarly the latter candidate can amend his policy so as to move into zone 2, thereby also securing support from foreign interests.

Policy combinations in zones 1 and 3 therefore are inconsistent with political equilibrium, and in zone 2 only the candidate announcing an export restraint has political support. Hence there exists no equilibrium characterized by a candidate's announcing a policy in terms of a tariff $t \in (0, \bar{t})$ when export restraints are available policy instruments.

Moreover, since $t = 0$ corresponds to $V = \tilde{V}$ and $t = \bar{t}$ to $V = 0$, tariffs are completely dominated by export restraints as policy instruments. In particular, this means that tariff equilibria which are not of the form $(t_0 = 0, t_1 = \bar{t})$ are not sustainable in the presence of the opportunity to use an export restraint.

It remains to demonstrate that an export-restraint equilibrium is unaffected by the option to use a tariff. To show this, begin with an equilibrium (V^*, V^*) and let one candidate deviate by announcing a tariff t' which attracts the support of domestic interests: so $\pi_i(t') > \pi_i(V^*)$. Proceed by introducing an export restraint V' such that domestic firms are indifferent between V' and t', that is, $\pi_i(V') = \pi_i(t')$. It can then be shown that $R(V^*, t') < R(V^*, V')$. However, since (V^*, V^*) is an equilibrium, $R(V^*, V') < R(V^*, V^*)$. Hence $R(V^*, t') < R(V^*, V^*)$, which proves that the deviating candidate cannot benefit. A candidate who deviates from (V^*, V^*) by announcing a tariff that attracts the support of foreign interests cannot benefit for similar reasons.

An important determinant of the result that export restraints drive out tariffs in the domestic political contest is the nonappropriability of tariff revenue by the political candidates. The candidates also cannot directly secure the rents associated with the restriction of trade via export restraints. However, these latter rents are a prize for foreigners and accordingly bear upon the value of campaign contributions received by the candidates. Thus, indirectly, the rents from restriction of trade under export restraints become part of campaign contributions and hence are of significance in the political contest. Tariff revenue has no corresponding political significance.

Campaign contributions made in the divergent tariff equilibrium will be greater than in the compromise export-restraint equilibrium. Indeed, since the export-restraint pronouncements converge in a Hotelling manner, there is nothing in equilibrium to distinguish the candidates. Strictly positive political outlays would however be made if the candidates kept part of the contributions for themselves, since then candidates can have an incentive not to converge completely in their policy pronouncements. A candidate's objective function would in that case encompass more than the probability of election. Introducing such a qualified objective function for political candidates does not change the comparative forms taken by the tariff and export-restraint equilibria. Tariff pronouncements still diverge and export-restraint pronouncements converge, but the latter convergence is less than complete.

V. Concluding Remarks

In prior analyses of the relation between political-influence motives and protection, a cohesive industry coalition has been set against a diffuse coalition of domestic losers from protection. Recognition of the domestic political-influence role of foreigners introduces another cohesive coalition to counter the domestic industry-specific interests seeking protection, and the tariff then becomes too divisive an instrument to be politically appealing. Our model of protectionist proposals as the equilibrium outcome of political competition has shown how tariffs are inconsistent with but export restraints facilitate the types of distributive compromises and opportunities for mutual collusive gain favored in political outcomes. Two questions remain.

First, if, as our model proposes, voluntary export restraints are the equilibrium outcome of political contestability of trade policy when tariffs can be used, why have export restraints become prominent only in recent years? We suggest that a political analogue to international integration of goods and financial markets has facilitated increased participation by foreign interests in domestic politics. Also, while there are certain legal complexities attendant on foreigners making domestic political outlays, campaign contributions are facilitated by the domestic legal entities which are subsidiaries of or have interests coincident with foreign corporations. Beyond these considerations, our model also suggests a role for changes in market structure. A more competitive foreign export industry has *ceteris paribus* more to gain from securing regulation of sales in its foreign markets. And a less competitive domestic import-competing industry has more to gain from supporting the restraint on domestic sales of its principal foreign competitors.

The second question is, given the prospect of collusive gain, why should foreign interests participate in the domestic political process rather than use their own home political process to secure a regulated restriction of exports? We would propose that such a collusive restriction of trade invites retaliation in export markets. The foreign cartel would be exposed as engaging in an unfair trade practice, in defiance of national trade laws and the General Agreement on Tariffs and Trade. However, this can clearly not be so when the export restraint is the outcome of the political process in the importing country which itself requests foreign exporters to cooperate in coordinating the restraint of trade.

REFERENCES

Baldwin, Robert, "The Political Economy of Protectionism," in *Import Competition and Response*, J. N. Bhagwati, ed., Chicago: University of Chicago Press, 1982, 263–86.

______, "Rent Seeking and Trade Policy: An Industry Approach," *Weltwirtschaftliches Archiv*, Heft 4, 1984, *120*, 662–77.

______, *The Political Economy of U.S. Import Policy*, Cambridge, MA: MIT Press, 1985.

Bernholz, Peter, "On the Reasons for the Influence of Interest Groups on Political Decision Making," *Zeitschrift für Wirtschafts und Socialwissenschaft*, 1974, *94*, 45–63.

______, "Dominant Interest Groups and Powerless Parties," *Kyklos*, 1977, *30*, 411–20.

Bhagwati, Jagdish N., "Protectionism: Old Wine in New Bottles," *Journal of Policy Modeling*, Spring 1985, *7*, 23–33.

______ **and Srinivasan, T. N.,** "Optimal Trade Policy and Compensation Under Endogenous Uncertainty: The Phenomenon of Market Disruption," *Journal of International Economics*, November 1976, *6*, 317–36.

Brecher, Richard A. and Bhagwati, Jagdish N., "Voluntary Export Restrictions Versus Import Restrictions: A Welfare-Theoretic Comparison," in H. Kierzkowski, ed., *Protection and Competition in International Trade: Essays in Honor of Max Corden*, Oxford: Basil Blackwell, 1987.

Cassing, James H. and Hillman, Arye L., "Political Influence Motives and the Choice Between Tariffs and Quotas," *Journal of International Economics*, November 1985, *19*, 279–90.

______ **and** ______, "Shifting Comparative Advantage and Senescent Industry Collapse," *American Economic Review*, June 1986, *76*, 516–23.

Collyns, Charles and Dunaway, Steven, "The Cost of Trade Restraints: The Case of Japanese Automobile Exports to the United States," *IMF Staff Papers*, March 1987, *34*, 150–75.

Cornes, Richard and Sandler, Todd, *The Theory of Externalities, Public Goods and Club Goods*, Cambridge: Cambridge University Press, 1986.

Das, Satya P., "Foreign Lobbying and the Political Economy of Protection," Indiana University, 1988.

Demsetz, Harold, "Industry Structure, Market Rivalry, and Public Policy," *Journal of Law and Economics*, April 1973, *16*, 1–10.

Eichenberger, Jurgen and Harper, Ian, "Price and Quantity Controls as Facilitating Devices," *Economics Letters*, 1987, *23*, 223–28.

Falvey, Rodney and Lloyd, Peter J., "The Choice of the Instrument of Industry Protection," in Richard N. Snape, ed., *Issues in World Trade Policy: GATT at the Crossroads*, New York: Macmillan, 1986, 152–70.

Feenstra, Robert C., "Voluntary Export Restraint in U.S. Autos 1980–1981: Quality, Employment and Welfare Effects," in Robert E. Baldwin and Anne O. Krueger, eds., *The Structure and Evolution of Recent U.S. Trade Policy*, Chicago: The University of Chicago Press for NBER, 1984, 298–325.

______, "Automobile Prices and Protection: The U.S.-Japan Trade Restraint," *Journal of Policy Modeling*, Spring 1985, *7*, 49–68.

Harris, Richard, "Why Voluntary Export Restraints Are Voluntary," *Canadian Journal of Economics*, November 1985, *18*, 799–809.

Hillman, Arye L., "Declining Industries and Political Support Protectionist Motives," *American Economic Review*, December 1982, *72*, 1180–87.

______, *The Political Economy of Protection*, in the series Fundamentals of Pure and Applied Economics, Chur, New York, London, Paris: Harwood Academic Publishers, 1988.

______ **and Katz, Eliakim** "Risk-averse Rent Seekers and the Social Cost of Monopoly Power," *Economic Journal*, March 1984, *94*, 104–10, reprinted in C. Rowley, R. Tollison, and G. Tullock, eds., *The Political Economy of Rent Seeking*, Boston: Kluwer Academic Publishers, 1988, 81–90.

______ **and Riley, John,** "Politically Contestable Rents and Transfers," presented at the conference on *Political Economy: Theory and Policy Implications*, World Bank, Washington D.C., June 1987, forthcoming in *Economics and Politics*, inaugural issue.

Husted, Steven, "Foreign Lobbying and the Formation of Domestic Trade Policy," presented at the Meetings of the Western Economic Association, San Francisco, July 1986.

Jones, Kent, "The Political Economy of Voluntary Export Restraints," *Kyklos*, 1984, *37*, 82–101.

Krishna, Kala, "Trade Restrictions as Facilitating Practices," *Journal of International Economics*, forthcoming 1988.

Mayer, Wolfgang, "Endogenous Tariff Formation," *American Economic Review*, December 1984, *74*, 970–85.

______ **and Riezman, Raymond,** "Endogenous Choice of Trade Policy Instruments," *Journal of International Economics*, November 1987, *23*, 377–81.

______ **and** ______, "Tariff Formation in

Political-Economy Models," presented at the conference on *Political Economy: Theory and Policy Implications*, World Bank, Washington D.C., June 1987.

Rodrik, Dani, "Tariffs, Subsidies and Welfare with Endogenous Policy," *Journal of International Economics*, November 1986, *21*, 285–99.

Young, Leslie and Magee, Stephen P., "Endogenous Protection, Factor Returns, and Resource Allocation," *Review of Economic Studies*, July 1986, *53*, 407–19.

Assisting Domestic Industries under International Oligopoly: The Relevance of the Nature of Competition to Optimal Policies

By LEONARD K. CHENG*

Optimal trade and industrial policies are derived for a home market that is supplied by a domestic firm and a foreign firm. The optimal policy combination can be quite sensitive to the nature of the duopoly's competition. For example, for some cost and demand structures, the optimal policy under Cournot competition consists of a domestic production tax and a tariff, but that under Bertrand competition consists of a production subsidy and free trade.

Until rather recently, economic analysis of international trade has been based almost exclusively on the assumption of perfectly competitive markets and occasionally on the opposite assumption of monopolies. Recognizing that markets often are neither perfectly competitive nor completely monopolized, trade theorists began several years ago to explore the implications of imperfect competition. Markets were modeled as international oligopolies comprised of firms with different nationalities,[1] and the following normative questions were frequently raised: (a) Does imperfect competition call for government intervention? and (b) What policies are optimal under imperfect competition?

While the existing work has provided many theoretical examples why the traditional analysis based on perfect competition may be misleading when markets are indeed imperfectly competitive, it has not come up with a general alternative to the traditional analysis. To a certain extent, this is to be expected because while there is an almost universally accepted notion of competitive equilibrium, there are at least several competing theories of oligopoly. It thus seems useful in the context of trade policy analysis to move in the direction of providing a general framework which includes widely accepted notions of oligopoly equilibrium as special cases.

From a normative point of view, it would be interesting to determine how optimal policies change with the nature of oligopolistic competition. If the differences between optimal policies for an oligopoly and those under perfect competition depend qualitatively on the details of firm behavior, then as far as policy analysis is concerned there does not exist a general alternative to the competitive analysis. That would indicate the importance of and a need for industry case studies which attempt to identify how firms in imperfectly competitive markets actually behave and interact in the real world. If instead the differences are qualitatively invariant to the exact details of the firms' interaction, then in a sense there is a general alternative to the traditional analysis. Even in this case, however, it would be interesting to determine how the magnitude of optimal policies varies with the nature of competition.

The paper by Jonathan Eaton and Gene Grossman, 1986, represents the first such systematic analysis for an export market. Depending on the nature of competition, they showed that the optimal policy could be an export tax, an export subsidy, or neither. When cost-reducing *R&D* is feasible, we can use similar arguments to show that the optimal policy varies from an export tax plus an *R&D* subsidy, to an export subsidy plus an *R&D* tax, to *laissez-faire* (i.e., free trade plus no *R&D* intervention).[2] Thus, for export markets at least, a general alternative to the competitive analysis does not exist because the optimal policies are qualitatively sensitive to the behavior of firms.

---

*Department of Economics, University of Florida, Gainesville, FL 32611. I am grateful to an anonymous referee for invaluable comments and particularly for the suggestion to imbed my earlier analysis in a general conjectural variations model. I am also indebted to Avinash Dixit, Gene Grossman, Gary Hufbauer, Richard Jensen, Richard Nelson, and Marie Thursby for their comments on earlier versions of this and related papers. Finally, support provided by the Center for International Economics and Business Studies at the University of Florida is greatly appreciated.

[1] For a recent survey and a systematic analysis of international oligopoly, see Avinash Dixit, 1984.

This paper, complementary to that of Eaton and Grossman, 1986, investigates how optimal policies toward domestic markets are affected by the nature of oligopolistic competition and the structure of demand. General results have been obtained for a wide range of oligopolistic behavior, but since Bertrand (price-setting) and Cournot (quantity-setting) competition are most widely studied in oligopoly theory and can be regarded as reasonable bounds on oligopolistic competition in the real world, we shall concentrate on these two forms of competition.

We start out with the case of differentiated products with linear demand functions. For any fixed degree of product differentiation, if the production costs of the domestic and foreign firms are not too different, then both domestic output and imports are desirable. An optimal policy would consist of a tariff and a domestic production subsidy. The production subsidy is used to remove domestic consumption distortion arising from the home firm's noncompetitive behavior, and the tariff is used to improve the home country's terms of trade. It turns out that the magnitude of optimal policies is smaller under Bertrand competition than under Cournot competition. The intuition is that firms are more competitive under the former so that there is less room for intervention.

If the firms' production costs are sufficiently different, then a corner solution will arise. We focus on the case where the foreign firm enjoys a large enough cost advantage so that the optimal level of domestic output is zero. As in the case of an interior solution, a tariff is used to improve the country's terms of trade, but now a production subsidy, by making the inactive home firm a credible threat to the foreign firm, also operates on the terms of trade indirectly. The tariff is still higher under Cournot competition, but the production subsidy is larger under Bertrand competition. The intuition is that a production subsidy is more efficient in improving the terms of trade under Bertrand competition, and as a result the role played by a tariff in that regard is diminished.

The degree of product differentiation has an important impact on the optimal policies. Under Bertrand competition, the tariff declines continuously as the substitutability between imports and domestic goods (call it k) increases; it approaches zero as the goods approach perfect substitutes. As k increases, the production subsidy declines to zero when the corner solution is first reached, but then rises again with further increases in k. The explanation is that when the goods become better substitutes, the production subsidy becomes more effective in improving the terms of trade, thus reducing the need for a tariff even though only imports are consumed.

The response of optimal policies to increases in k is quite different under Cournot competition. The tariff declines initially but then rises again, reaching the level which corresponds to zero substitutability when k is large enough to yield the corner solution! The behavior of the production subsidy is similar to that under Bertrand competition before the corner solution is first reached. After that it is no longer unique. It may be positive and rising with k, but it can as well be zero or negative. The fact that the foreign firm is not restrained at all by the home firm under Cournot competition when the latter does not produce explains why the tariff rises back up as domestic output declines to zero.

Finally, we relax the assumption of linear demand but assume that imports and domestic goods are perfect substitutes, which implies that the optimal solution is necessarily corner under Bertrand competition. If domestic production costs exceed foreign production costs, then the optimal policy under Bertrand competition consists of free trade and a positive production subsidy, irrespective of the shape of the demand curve.

---

[2] See Leonard Cheng, 1987.

Nonlinear demand implies that a production subsidy can be used to improve the terms of trade when the optimal solution under Cournot competition is interior, something that could not happen with linear demand. In addition, depending on the curvature of the demand curve, one of the policy instruments may be negative. For example, if the demand curve is sufficiently concave and if the cost advantage of the foreign firm is sufficiently large so that the amount of domestic output is small, the optimal policy under Cournot competition consists of a tariff and a production tax. This is in sharp contrast to the optimal policy under Bertrand competition.

Section I briefly reviews the literature and Section II develops a general conjectural variations model. Optimal policies are derived for the case of differentiated products with linear demand functions in Section III and for the case of homogeneous goods with a general demand function in Section IV. Section V offers a brief summary and some concluding remarks. For completeness, the case of differentiated products with general demand functions is contained in the Appendix.

I. A Brief Review of the Literature

Eaton and Grossman, 1986, have pointed out that optimal policies toward export markets are sensitive to the nature of oligopolistic competition.[3,4] It has been shown by them and others that an export subsidy is likely to be indicated under Cournot competition if the number of exporting firms is small.[5] In contrast, an export tax is likely to be optimal under Bertrand competition.[6] In both cases, if firms have consistent conjectures in equilibrium (i.e., a firm's conjectural variations about other firms are equal to those firms' optimal response to the firm's action),[7] then neither an export tax nor an export subsidy is called for when there is one domestic exporter, but an export tax is optimal if there are two or more domestic exporters.

In the three-stage game of industrial strategy pioneered by Spencer and Brander, 1983, the governments of the domestic and foreign exporters (assumed to be one in each country) set *R&D* and trade policies before the firms incur *R&D* expenditure (to lower their respective production costs) and set their respective output levels. Assuming that the market equilibrium was Cournot, they showed that the optimal policy of each government consisted of an *R&D* tax plus an export subsidy. Using a parallel argument, however, one can show that if the market equilibrium is Bertrand, then the optimal policy of each government would consist of an *R&D* subsidy plus an export tax.[8] Furthermore, if the firms have consistent conjectures, then *laissez-faire* (i.e., free trade plus no *R&D* intervention) is optimal.[9]

---

[3] In the literature, export markets are often analyzed under the assumption that they are independent of domestic markets or even more strongly that there is no domestic consumption.

[4] Related work was done by Kala Krishna, who studied the effect of different forms of trade policies on oligopolistic competition. Tariffs and quotas were compared in Krishna, 1985, while physical and value-added content protection was the focus of Krishna and Motoshige Itoh, 1988. In neither paper, however, were optimal policies derived.

[5] See also James Brander and Barbara Spencer, 1985; and Dixit, 1984.

[6] See also Cheng, 1984.

[7] The notion of consistent conjectures equilibrium was discussed in Timothy Bresnahan, 1981, and Martin Perry, 1982, among others. It was motivated mainly by the hope that consistency requirements would lead to a unified theory of oligopoly by ruling out equilibria that are inconsistent. Unfortunately, the hope was not realized. Marcel Boyer and Michel Moreaux, 1983, demonstrated that, with appropriate conjectural functions, any sustainable output combination can be obtained as a locally consistent conjectures equilibrium. Partly as a result of their demonstration, and partly due to the argument that conjectures are not meaningful in a static game, the notion of consistent conjectures equilibrium is quite unpopular among game theorists.

[8] See Cheng, 1987.

[9] See Cheng in the work cited.

Unlike the case of export markets, less clear-cut results have been obtained about optimal policies toward domestic markets. A major reason is that different researchers have looked at different policy instruments. For instance, Brander and Spencer, 1984, focused on a tariff (or import subsidy) and Eaton and Grossman, 1986, focused on a production subsidy (or tax), whereas Cheng, 1984, and Dixit, 1984, 1988, considered both a tariff (or import subsidy) and a domestic production subsidy (or tax).

Assuming that firms produce a homogeneous product and the market equilibrium is Cournot, Brander and Spencer, 1984, showed that a tariff was desirable if it did not worsen the home country's terms of trade too much relative to the domestic firms' increase in market share.[10] Within a similar framework, Dixit, 1984, showed that if a country had a cost advantage over imports, the optimal policies would be a prohibitive tariff plus a domestic subsidy.[11] On the other hand, if the home firms had a cost disadvantage which was not too great, then both imports and domestic production would be desirable. In that case, a tariff would be optimal if demand was linear.

In contrast with Brander and Spencer, 1984; and Dixit, 1988, Eaton and Grossman, 1986, assumed that a homogeneous good was sold at the same price in domestic and foreign markets due to perfect arbitrage and that the market equilibrium was given by a consistent conjectures equilibrium in which quantity is the strategy variable. They showed that a domestic production subsidy was optimal if the foreign firms' marginal cost was rising. However, if the firms' marginal costs were constant, then neither a production tax nor a production subsidy would be optimal.[12]

Unfortunately, because different policies or policy combinations were considered in the above-mentioned studies, it is difficult to see from these results whether and how optimal policies depend on the nature of oligopolistic competition. To address the question directly, one must consider the same set of policies under different forms of competition. That will be the task of the next three sections.

II. The Model

I shall construct a general conjectural variations model of a duopoly consisting of a domestic firm and a foreign firm. In this paper, conjectural variations are used to *represent* different forms of oligopolistic competition resulting from the adoption of different strategies, such as Bertrand and Cournot competition, rather than to imply that firms actually attempt to anticipate how their rivals would *react* to their own quantities of supply.[13] To highlight optimal policies in the presence of domestic consumption, I assume that the domestic firm does not export at all. Finally, as in Dixit, 1984, 1988, and Cheng, 1984, I consider the joint use of a tariff (or import subsidy) and a domestic production subsidy (or tax).

Consider two firms, one domestic and one foreign, producing two substitutes for a home market. Following Dixit, 1984, and Eaton and Grossman, 1986, I adopt the convention that lowercase letters denote variables associated with the domestic firm and uppercase letters denote variables associated with the foreign firm. Demand is assumed to be generated by a representative consumer whose utility function is both separable and linear in a numeraire good. Let V be the consumer's surplus from the consumption of the goods.

---

[10] They assumed that there were no export markets and the domestic market was served by a foreign firm and a domestic firm.

[11] Dixit assumed that the domestic market was independent of export markets but allowed for many domestic and foreign firms. After this paper was completed, I learned of a related paper by K. C. Fung, 1987, who studied five alternative policies (namely, tariffs, production subsidies, consumption taxes, countervailing duties, and antidumping duties) in a differentiated duopoly with linear demand and Cournot conjectures.

[12] When the firms' marginal costs are constant and equal, the consistent conjectures equilibrium becomes a Bertrand or competitive equilibrium, that is, price is equal to marginal costs.

[13] See fn. 7.

By definition,

$$(1)\quad V(p,P) \equiv \max_{q,Q} U(q,Q) - pq - PQ,$$

where $q[p]$ is the quantity (price) of the good produced by the home firm, $Q[P]$ is the quantity (price) of the good produced by the foreign firm, and U is measured in terms of the numeraire.

The inverse demand functions are given by

$$(2)\qquad p = p(q,Q),$$

$$(3)\qquad P = P(q,Q),$$

where $p = \partial U/\partial q$ and $P = \partial U/\partial Q$. As the goods are substitutes, the inverse demand functions satisfy the following conditions

$$(4)\qquad p_1 \leqq p_2 < 0; \quad P_2 \leqq P_1 < 0,$$

where $p_i[P_i]$ is the derivative of $p[P]$ with respect to its ith argument, $i = 1,2$. Because $\partial^2 U/\partial q \partial Q = \partial^2 U/\partial Q \partial q$, $p_2 = P_1$. If the goods are homogeneous, then $p \equiv P$ and $p_1 = p_2$.

Let t denote a specific tariff on imports (t becomes a subsidy if it is negative), and s denote a unit subsidy for the home firm's output (s becomes a tax if it is negative). National welfare in the domestic market is equal to the sum of (a) the consumer's surplus, (b) the home firm's profits, and (c) the tariff revenues, less (d) production subsidies, that is,

$$(5)\qquad \begin{aligned} W &= (U(q,Q) - pq - PQ) \\ &\quad + (p - c + s)q + tQ - sq \\ &= U(q,Q) - cq - (P - t)Q, \end{aligned}$$

where c is the home firm's constant marginal costs,[14] and q and Q are determined by a Nash equilibrium of the following conjectural variations model:

$$(6)\qquad p(q,Q) + q(p_1 + p_2\gamma) - c + s \leqq 0,$$

$$(7)\quad P(q,Q) + Q(P_2 + P_1\Gamma) - C - t \leqq 0,$$

where C is the foreign firm's constant marginal costs, and $\gamma \equiv dQ/dq$ and $\Gamma \equiv dq/dQ$ are the domestic and foreign firms' conjectural variations, respectively. The strict inequality in (6) ((7)) may hold only if $q[Q]$ is equal to zero.

By definition, $\gamma = \Gamma = 0$ under Cournot competition. In contrast, $\gamma = -P_1/P_2$ and $\Gamma = -p_2/p_1$ under Bertrand competition.[15] Since our purpose is to examine the dependence of optimal policies on the nature of oligopolistic competition, we would like to treat γ and Γ as parameters that can be varied exogenously. This implies that, for the conjectural variations model to retain Bertrand competition as a special case, the demand functions must be such that both P_1/P_2 and p_2/p_1 are independent of q and Q, or equivalently q and Q must enter the functions as a weighted sum.[16] In the main body of this paper, I shall deal with two special cases which possess this property, namely, the case of differentiated products with linear demand functions and the case of homogeneous goods with a general demand function.

The welfare effect of an infinitesimally small change in the Nash equilibrium is given by the total differential of (5),

$$(8)\quad dW = (p - c)\,dq + t\,dQ - Q\,d(P - t),$$

where dq, dQ, and $d(P - t)$ represent infinitesimal changes brought about by an infinitesimal change in parameters. The first term captures the domestic consumption distortion effect: an increase in q (other things held constant) increases welfare if $p > c$. The second term captures the import consumption distortion effect:[17] an increase in Q (other things held constant) increases welfare if $t > 0$, because while consumers pay P for each unit of imports the net cost is only $(P - t)$. The last term captures the terms-of-trade effect. These welfare components will be useful in explaining our subsequent results.

---

[14] For an analysis of inter-market linkages under oligopoly due to nonconstant marginal costs, see Jeremy Bulow, John Geanakoplos, and Paul Klemperer, 1985; and Paul Krugman, 1984.

[15] Under Bertrand competition, the home firm regards P as given. Since $dP \equiv P_1 dq + P_2 dQ$ by definition, it follows from $dP = 0$ that $\gamma \equiv dQ/dq = -P_1/P_2$. Similarly, $\Gamma = -p_2/p_1$ because the foreign firm regards p as given.

[16] This follows as a special case of Jànos Aczél's 1985 theorem.

Unless stated to the contrary, I assume henceforth that (6) and (7) hold with equality so that there is an effective duopoly at the optimal solution. As long as $[p + q(p_1 + p_2\gamma)]$ and $[P + Q(P_2 + P_1\Gamma)]$ are not linearly dependent, q and Q are determined uniquely by s and t. Substituting (7) into (5), national welfare can be written as

$$W = U(q,Q) - cq - [C - Q(P_2 + P_1\Gamma)]Q, \tag{9}$$

where $[C - Q(P_2 + P_1\Gamma)]$ is the unit price of imports net of the tariff, that is, $(P - t)$.

The home government's problem is to choose s and t, which maximize (5) subject to (6) and (7). But, as Dixit, 1984, 1988, has amply demonstrated, it is analytically more fruitful to maximize (9) with respect to q and Q. The analysis to be presented in the next two sections is structurally very similar to that of Dixit, 1988, but there are important differences.[18]

III. Differentiated Products with Linear Demand Functions

Let the inverse demand functions be given by

$$p = a - bq - kQ, \tag{10}$$

$$P = A - kq - BQ, \tag{11}$$

where all the parameters are positive and $bB - k^2 > 0$. I assume that both $(b + k\gamma)$ and $(B + k\Gamma)$ are nonnegative. This assumption is hardly restrictive at all because the admissible values of γ and Γ include as special cases perfect competition ($\gamma = -b/k$, $\Gamma = -B/k$), Bertrand competition ($\gamma = -k/B, \Gamma = -k/b$), Cournot competition ($\gamma = \Gamma = 0$), and perfect collusion ($\gamma = Q^m/q^m, \Gamma = 1/\gamma$, where q^m and Q^m maximize the firms' joint profits).

Using (7) and (11) in (8), we have

$$dW = (p - c)\,dq + [t - Q(B + k\Gamma)]\,dQ. \tag{12}$$

Thus, if it is optimal to consume domestic goods, then $p = c$ and (from (6) and (10)) $s^* = q(b + k\gamma)$, which is strictly positive unless the home firm is perfectly competitive. The optimal subsidy is used solely to remove the domestic consumption distortion arising from the home firm's noncompetitive behavior. It will be shown in later sections that s^* also plays a role in affecting the terms of trade if demand is nonlinear or if domestic output is zero and Γ is negative.

If it is optimal to have imports, then the optimal tariff t^* is equal to $Q(B + k\Gamma)$. It is strictly positive unless the foreign firm is perfectly competitive. The optimal tariff is used to restrict imports until the marginal benefit from an improvement in the country's terms of trade is just offset by the marginal cost of import consumption distortion.

A. *Domestic Goods and Imports Are Both Desirable*

Combining (6) and (7) with (10) and (11), one can solve explicitly for the quantities and prices (see Dixit's, 1988, equations (15) and (16) or Cheng, 1987, for details). Substituting these quantities into s^* and t^*, one can express the optimal policies explicitly in terms of the firms' conjectural variations and other parameters. For future references, the optimal policies under Bertrand and Cournot competition (denoted by subscripts B and C, respectively) are as follows when both q and Q are positive:

---

[17] This was interpreted as the volume-of-trade effect in Dixit, 1988.

[18] Both papers examine the comparative static response of optimal tariffs and subsidies to the conjectural variation parameter, but they are different in a number of aspects. First, our results are interpreted as the relevance of the nature of competition to optimal policies, whereas Dixit's were interpreted as anti-dumping duties and countervailing subsidies. Second, this paper focuses on a duopoly while Dixit allows for a large number of identical firms in each country. Third, we give special attention to corner solutions and the cases of Cournot and Bertrand competition. Fourth, optimal policies at corner colutions obtained here for the case of homogeneous goods are markedly different from those of Dixit's (see fn. 21 below).

$$t_B^* = \frac{1}{3bB}\left[bB(A-C) - kB(a-c)\right], \tag{13}$$

$$s_B^* = \frac{1}{3bB}\left[(3bB - 2K^2)(a-c) - bk(A-C)\right], \tag{14}$$

$$t_C^* = \frac{B}{(3bB - k^2)}\left[b(A-C) - k(a-c)\right], \tag{15}$$

$$s_C^* = \frac{b}{(3bB - k^2)}\left[3B(a-c) - k(A-C)\right]. \tag{16}$$

It is straightforward to show that $t_C^* > t_B^*$ and $s_C^* > s_B^*$. Indeed, one can obtain a more general result by taking the derivatives of s^* and t^* with respect to γ and Γ.

$$\frac{dt^*}{d\Gamma} = kQ\Delta/\Delta'' \geqq 0, \tag{17}$$

$$\frac{dt^*}{d\gamma} = 0, \tag{18}$$

$$\frac{ds^*}{d\Gamma} = 2k^2(b + k\gamma)Q/\Delta'' \geqq 0, \tag{19}$$

$$\frac{ds^*}{d\gamma} = kq \geqq 0, \tag{20}$$

where $\Delta'' \equiv [(3B + 2k\Gamma)b - k^2] \geqq \Delta \equiv bB - k^2 > 0$.

Recalling that an increase in competitiveness on the part of the firms is reflected by a decrease in γ and Γ, we conclude from equations (17)–(20) that the magnitude of t^* and s^* declines as the firms become more competitive. This can be explained by the fact that there are fewer distortions to offset if the firms are more competitive.

Taking derivatives of q and Q with respect to γ and Γ and making use of (17)–(20), one can show that $\partial q/\partial\gamma = \partial Q/\partial\gamma = 0$, $\partial Q/\partial\Gamma < 0$, and $\partial q/\partial\Gamma > 0$. The first two results can be explained by the fact that the home firm's noncompetitive behavior (as reflected by γ) is completely offset with an optimal subsidy (recall the discussion immediately below (12)), so that the social costs of home goods are independent of γ. In contrast, the social unit cost of imports, $[C + Q(B + k\Gamma)]$, depends on Γ. It is intuitive that when the foreign firm is more competitive (i.e., Γ falls), imports are more affordable and thus more of them should be consumed. Since domestic goods and imports are substitutes, any increase in imports implies less domestic output.

It is perhaps useful to point out that if demand is symmetric (i.e., $a = A$, $b = B$) and the firms have identical production costs (i.e., $c = C$) and behave identically (i.e., $\gamma = \Gamma$), then $q > Q$. This is a reflection of the fact that the social unit cost of home goods is c but that of imports exceeds C. Thus, there is a bias in favor of domestic goods.

B. *Corner Solutions*

The case of $Q = 0$ and $q > 0$ arises if $a > c$ and $A - k(a-c)/b \leqq C$. The optimal policies are easy to characterize. A production subsidy is used to ensure that $p = c$. No tariffs are necessary, but any positive tariff will do just as well because it does not matter whether the duopoly degenerates into a domestic monopoly.

The case of $q = 0$ and $Q > 0$ is analytically more interesting. The conditions under which this case arises are $A > C$ and $a - kQ \leqq c$, where Q is the optimal level of imports. Since $t = Q(B + k\Gamma)$, it follows from (7) that $P = C + 2Q(B + k\Gamma)$. Together with (11) and the fact that $q = 0$ this implies that $Q =$

$(A-C)/(3B+2k\Gamma)$. As a result, the second condition becomes

$$(21) \qquad (a-c) \leqq \frac{k(A-C)}{(3B+2k\Gamma)}.$$

Thus, in order for the market to be supplied with imports only, the foreign firm must enjoy an advantage which is sufficiently great, that is, $(A-C)$ must exceed $(a-c)$ by a large enough margin. The required margin becomes smaller if the foreign firm is more competitive (i.e., Γ declines) or if the goods are better substitutes (i.e., k rises). In the limiting case, if the foreign firm is perfectly competitive, then the required margin becomes arbitrarily small as the goods approach perfect substitutes (which implies that k approaches B).

The optimal tariff is given by

$$(22) \quad t^* = Q(B+k\Gamma)$$

$$= (A-C)(B+k\Gamma)/(3B+2k\Gamma),$$

from which it follows that $t_B^* = (bB-k^2)(A-C)/(3bB-2k^2)$ and $t_C^* = (A-C)/3$. As in the case of an interior solution, $t_C^* > t_B^*$.[19] In addition, t_B^* is close to zero if the goods are extremely close substitutes (i.e., $k^2 \rightarrow bB$).[20]

When $q=0$, the optimal production subsidy is obviously not used to reduce domestic consumption distortion. Instead, it is used mainly to improve the terms of trade indirectly by making the inactive home firm a credible threat to the foreign firm. More precisely, in order for the foreign firm to maintain any negative Γ (such as under Bertrand competition) when the home firm is not producing at all, it requires the home firm to be right on the market's edge, that is, (6) holds with equality at the optimal solution with $q=0$ so that the two firms constitute an effective duopoly. From this requirement we obtain

$$(23) \qquad s^* = \frac{k(A-C)}{(3B+2k\Gamma)} - (a-c),$$

and as a special case $s_B^* = [bk(A-C)/(3bB-2k^2)] - (a-c)$.

Upon comparison with (21), one sees that $s^* = 0$ if (21) holds as an equality, but $s^* > 0$ if (21) holds as a strict inequality. In addition, s^* increases with c but decreases when C goes up. Finally, s_B^* is arbitrarily close to $(c-C)$ if the goods become extremely close substitutes (i.e., $a \rightarrow A$, $b \rightarrow B$ and $k \rightarrow B$).[21]

If $\Gamma \geqq 0$, it is not necessary for (6) to hold as an equality. Consequently, the equality sign in (23) can be replaced by a weak equality sign ($\leqq$). Under Cournot competition, $\Gamma = 0$ and thus $s_C^* \leqq [k(A-C)/3B] - (a-c)$. In contrast with the interior solution, we now see that $s_B^* > s_C^*$.

C. *The Degree of Product Differentiation*

An increase in k toward b and B reflects an increase in substitutability between home goods and imports. At one extreme the goods are completely independent if $k=0$. At the other extreme the goods are perfect substitutes if $k=b=B$ and $a=A$. To explore how the degree of product differentiation interacts with firm behavior in determining optimal policies, let us focus on Cournot and Bertrand competition.

The optimal policies under Bertrand and Cournot competition were given above in

---

[19] More generally, it can be established that $\partial t^*/\partial \Gamma > 0$.

[20] See Section III, Part C below.

[21] Thus, under Bertrand competition with extremely close (but nonetheless imperfect) substitutes, t^* is arbitrarily close to zero and s^* is arbitrarily close to $(c-C) > 0$. This finding corroborates our results in the next section, but is markedly different from that implicit in Dixit (1988, p. 66): s^* is arbitrarily close to zero and t^* is arbitrarily close to $(c-C)$ if the goods are homogeneous and competition is sufficiently close to Bertrand. The discrepancy in findings is attributable to the use of different methods of derivation. If our present method were used instead in Dixit's model for the case of homogeneous goods (i.e., expressing dW as a function of dq and dQ rather than of dQ and $d(q+Q)$) when $q=0$, then his finding would have been identical to ours.

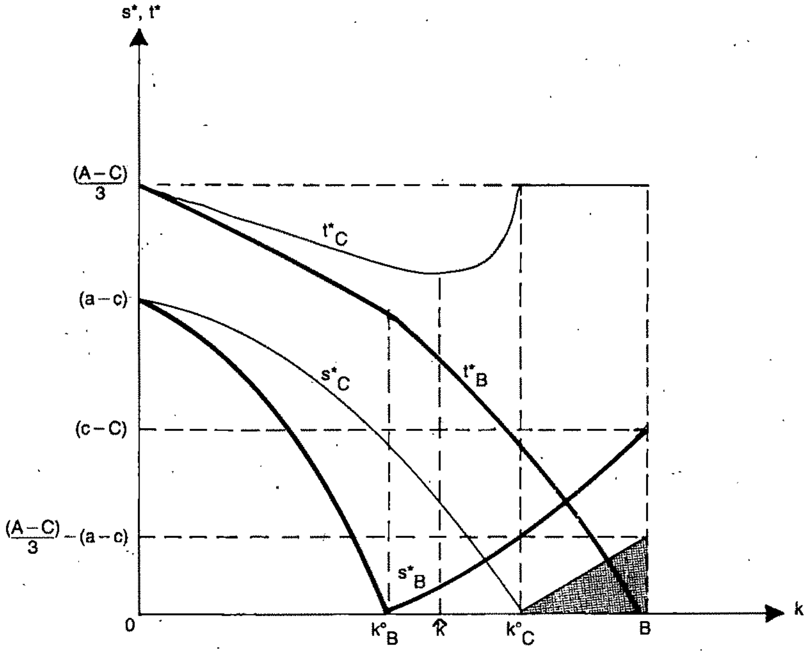


FIGURE 1. OPTIMAL POLICIES AND THE DEGREE OF PRODUCT DIFFERENTIATION

equations (13)–(16) for an interior solution and beneath equations (22) and (23) for the corner solution at which $q=0$ and $Q>0$. The condition under which an interior solution gives way to the corner solution is $bk(A-C)/(3bB-2k^2) \geq (a-c)$ under Bertrand competition and $k(A-C) \geq 3B(a-c)$ under Cournot competition. In either case, the condition is more likely to be met when k rises. To ensure that the optimal solution can be either interior or corner depending on the value of k, we assumed that $(A-C) \gg 3(a-c) > 0$. This condition implies that there exist two numbers, k_B° and k_C°, satisfying $0 < k_B^\circ < k_C^\circ < B$ such that the corner solution prevails if and only if $k \geqq k_B^\circ$ ($k \geqq k_C^\circ$) under Bertrand (Cournot) competition.

The relationship between optimal policies under Bertrand and Cournot competition and the degree of product differentiation is summarized graphically in Figure 1, where we have assume that $a \to A$ and $b \to B$ as $k \to B$. It should come with no surprise that $t_C^* = t_B^*$ and $s_C^* = s_B^*$ at $k=0$ because the duopoly is then comprised of two independent monopolies. As k increases, the firms continue to exert greater competitive pressure on each other, leading to a decline in the magnitude of optimal policies. At k_B° s_B^* is equal to zero as q_B first becomes zero. But as k increases beyond k_B°, s_B^* rises with k while t_B^* continues to fall. In the limit, when the goods become almost perfect substitutes, s_B^* rises to near $(c-C)$ and t_B^* falls to near zero.

In contrast with t_B^*, t_C^* declines as k rises until $\hat{k}$ is reached, from which point onward it begins to rise. At k_C° (when q_C first becomes zero) and beyond, the tariff is equal to that at $k=0$! The production subsidy s_C^*, on the other hand, falls as k increases until k_C° is reached. In the range (k_C°, B) s_C^* is nonunique and any point in the shaded area is optimal.

IV. Homogeneous Products

When home goods and imports are perfect substitutes, p and P coincide and q and Q

enter the common inverse demand function as a simple sum, that is, $p \equiv P = D(q+Q)$. The change in welfare can be written as

$$(24)\quad dW = [p - c + Q^2(1+\Gamma)D'']\,dq + [p - C + Q(1+\Gamma)(2D' + QD'')]\,dQ,$$

where D' and D'' are the first- and second-order derivatives of D, respectively, and use has been made of (7). As q is increased by an infinitesimal amount, welfare changes at a rate of $(p-c)$ due to domestic consumption distortion arising from the home firm's noncompetitive behavior, and at an additional rate of $(Q^2(1+\Gamma)D'')$ through the terms of trade. In a similar way, as Q is increased by an infinitesimal amount, welfare changes at a rate of $(p - C + Q(1+\Gamma)D') = t$ due to import consumption distortion in the presence of a tariff (or import subsidy), and at an additional rate of $[Q(1+\Gamma)D' + Q^2(1+\Gamma)D'']$ through the terms of trade. The optimal policies to be derived below reflect these distortions.

If it is optimal to produce goods at home, then $p - c + Q^2(1+\Gamma)D'' = 0$. If it is optimal to import, then $p - C + Q^2(1+\Gamma)D'' + 2Q(1+\Gamma)D' = 0$. Since the term $2Q(1+\Gamma)D'$ is negative for any positive Q unless the firms are perfectly competitive,[22] it follows that there should be no imports if $c \leqq C$. On the other hand, if $c > C$, then some imports are always desirable as long as home goods are produced. If both q and Q are positive, then $[c - C + 2Q(1+\Gamma)D'] = 0$.

A. *Interior Solutions*

Combining the first-order conditions with (6) and (7), we obtain the optimal policies when both q and Q are positive.[23]

$$(25)\quad s^* = Q^2(1+\Gamma)D'' - q(1+\gamma)D',$$

$$(26)\quad t^* = -Q(1+\Gamma)(QD'' + D').$$

Recalling that $Q^2(1+\Gamma)D''$ captures the terms of trade effect, we see that with nonlinear demand, the optimal domestic subsidy is used to improve the country's terms of trade even when both goods are consumed. If the inverse demand function is convex, then s^* is greater than that which completely removes domestic consumption distortion. If the demand function is concave, the s^* is smaller than that which completely removes domestic consumption distortion, which implies that s^* may be negative. Similarly, t^* may be negative if the inverse demand function is highly convex.

B. *Corner Solutions*

We now turn to corner solutions in general and to the case of Bertrand competition in particular. Again, we focus on the more interesting case of $q = 0$, which will arise if $D'(0) > C$ and $c - C \geqq -2Q(1+\Gamma)D'$, where D' is evaluated at Q. The optimal tariff is still given by (26) with q set equal to zero, but $s^* \leqq c - p$, where the strict inequality sign holds only if $\Gamma \geqq 0$.

When the firms produce homogeneous goods and have unequal constant marginal costs (i.e., $(C+t) \neq (c-s)$), one can think of Bertrand competition in pure strategies resulting in an equilibrium in which the market is captured by whichever firm with the lower cost and the equilibrium price is just below the other firm's marginal cost,[24] provided that the high-cost firm's marginal cost does not exceed the low-cost firm's monopoly price.[25] For instance, if $(C+t) < (c-s)$, then the equilibrium price is given by $(c-s-\varepsilon)$ and the domestic market is supplied with imports only, where ε is an arbitrarily small but strictly positive number.

---

[22] While obvious, it may be useful to point out the difference in firm behavior between perfect competition and Bertrand competition with homogeneous goods. Under perfect competition, a firm supplies according to its marginal cost curve, but under Bertrand competition each firm tries to set its own price to maximize its own profits subject to prices set by rival firms.

[23] One can see from the first-order conditions that q and Q depend on Γ but not on γ. In conjunction with (26), we conclude that $dt^*/d\gamma = 0$, which is identical to (18) in Section III. This result can be explained again by the fact that the social costs of home goods and imports are independent of γ.

[24] One can also think of Bertrand competition in mixed strategies, as in Eric Fisher and Charles A. Wilson, 1987. When firms compete in one another's domestic market under the condition that each firm has to charge a single FOB price for both markets, they have shown that a tariff serves to increase the profits of both firms. They have not, however, considered the question of optimal policies.

Let us consider the optimal policy under Bertrand competition for the case in which $c > C$. Obviously, no domestic output is desirable if the marginal social cost of imports is less than c, which can be achieved by subsidizing the home firm until $(c-s)$ is just above $(C+t)$. Given the optimal production subsidy, a zero tariff is optimal because any tariff or import subsidy will only distort consumption. Therefore, $s_B^* = c - C - \varepsilon$ and $t_B^* = 0$, which is basically identical to the results obtained in Section III, Part B, for the case of Bertrand competition with extremely close substitutes. This is to be expected, since homogeneous goods are a limiting case of close substitutes. More important than providing a consistency check, however, this example illustrates the rationale for subsidizing home firms to keep potential competitive pressure on foreign firms.

C. *Optimal Policies under Bertrand and Cournot Competition*

From (25) and (26) one sees that if c exceeds C by a sufficient amount so that q is small, then $t_C^* > 0$ and $s_C^* < 0$ if the demand curve is sufficiently concave. In sharp contrast, $t_B^* = 0$ and $s_B^* > 0$ if $c > C$, irrespective of the shape of the demand curve. Thus, with the same demand and cost structure, optimal policies under the two different forms of oligopolistic competition can be qualitatively quite different.

To explain the above results, it is useful to recognize the fact that under Cournot competition the impact of the domestic firm on the foreign firm's actions is directly proportional to domestic output. In contrast, under Bertrand competition the domestic firm could have a significant bearing on the foreign firm without actually supplying a large quantity to the market, provided that it is ready to supply a large quantity if imports are priced above certain levels. Thus, the social costs of a domestic production subsidy, relative to those of a tariff or import subsidy, would be lower under Bertrand competition than under Cournot competition. Moreover, when the goods become better substitutes, the potential restraining power of the domestic firm on the foreign firm increases and the difference in social costs would increase. The case of homogeneous goods brings out the result most clearly, since no domestic subsidy is actually paid out as the domestic firm never actually produces.

V. Summary and Concluding Remarks

We have considered a domestic market that is served by a domestic firm and a foreign firm. Optimal trade and industrial policies were derived in a general conjectural variations model, but special attention was given to Bertrand and Cournot competition. We studied the dependence of optimal policies on the nature of oligopolistic competition and the structure of demand. In the case of differentiated products with linear demand functions, if the goods are sufficiently differentiated and the cost advantage of the foreign firm is not too large, then both imports and domestic goods are consumed. The optimal policy under either Cournot or Bertrand competition consists of a tariff and a domestic production subsidy, but the magnitude of policy is smaller under the latter form of competition. If the goods are sufficiently close substitutes or if the cost advantage of the foreign firm is sufficiently large, then only imports will be consumed. While the optimal policy under Bertrand competition consists of a production subsidy, that under Cournot competition may consist of a production tax or subsidy. Thus, the optimal policies may be qualitatively different at corner solutions. If we are willing to ignore the (equally) optimal production taxes in the Cournot case, then we can say that they are qualitatively similar. But even in that case the magnitude of tariff is smaller and that of production subsidy larger under Bertrand competition.

---

[25] The equilibrium price is equal to the low-cost firm's monopoly price if the condition is not satisfied. In the presence of optimal policies, however, the condition will always be satisfied in our subsequent analysis.

In the case of perfect substitutes with a general demand function, optimal policies are more likely to be qualitatively different. For instance, if domestic production costs exceed foreign production costs, then the optimal policy under Bertrand competition consists of a production subsidy and free trade, irrespective of the shape of the demand curve. The optimal policy under Cournot competition, on the other hand, depends importantly on the demand curve's curvature. If both imports and domestic goods are consumed, it may consist of a tariff and a production tax, or an import subsidy and a production subsidy. If only imports are consumed, then virtually any combination can be optimal.

Since linear demand is only a very special case, we must conclude that optimal trade and industrial policies are qualitatively sensitive to the nature of competition among firms, and thus a general alternative to competitive analysis does not exist. The fundamental reason is the abundance of oligopoly equilibrium solutions resulting from the possible use of different strategies. This points to the importance of and a need for industry case studies in future research. Unless and until one can identify, perhaps approximately, the nature of oligopolistic interaction in the real world (which of course can vary across industries and markets), one is forced to be skeptical about certain policy recommendations emerging from the theoretical analysis of international oligopoly.

APPENDIX: OPTIMAL POLICIES UNDER BERTRAND AND COURNOT COMPETITION WITH GENERAL DEMAND FUNCTIONS

Using the same notation as in Section II and assuming that both q and Q are positive, one can derive the following optimal policies:

$$\text{(A1)} \quad s_C^* = Q_C^2 P_{21} - q_C p_1,$$

$$\text{(A2)} \quad t_C^* = -Q_C(P_2 + Q_C P_{22}),$$

$$\text{(A3)} \quad s_B^* = Q_B^2\left[P_{21} - P_{11}\frac{p_2}{p_1} - P_1\left(\frac{p_1 p_{21} - p_2 p_{12}}{p_1^2}\right)\right] - q_B\left(p_1 - p_2\frac{P_1}{P_2}\right),$$

$$\text{(A4)} \quad t_B^* = -Q_B\left[P_2 - P_1\frac{p_2}{p_1} + Q_B\left(P_{22} - P_{12}\frac{p_2}{p_1} - P_1\left(\frac{p_1 p_{21} - p_2 p_{12}}{p_1^2}\right)\right)\right],$$

where B and C stand for Bertrand and Cournot competition, respectively.

The optimal policies under Bertrand competition can alternatively be expressed in terms of direct demand functions. Denoting the direct demand functions invertible from $p(q, Q)$ and $P(q, Q)$ by $q(p, P)$ and $Q(p, P)$, we have

$$\text{(A5)} \quad s_B^* = \frac{Q^2[Q_1 Q_{22} - Q_2 Q_{21}]}{\delta Q_2^2} - \frac{q}{q_1},$$

$$\text{(A6)} \quad t_B^* = \frac{Q\left[q_2(Q_2 Q_1 - Q Q_{21}) - q_1(Q_2^2 - Q Q_{22})\right]}{\delta Q_2^2},$$

where Q_i and q_i are the derivatives of Q and q with respect to their ith arguments, respectively, $i = 1, 2$, and Q_{2j} is the derivative of Q_2 with respect to its jth argument, $j = 1, 2$. By definition, $\delta \equiv (Q_2 q_1 - Q_1 q_2)$ is positive if the goods are imperfect substitutes, that is, $p_1 P_2 > p_2 P_1$.

Based upon the expression for the net price of imports and the above formulas, one sees that t_C^* and t_B^* are used to improve the home country's terms of trade. In contrast, s_C^* as well as s_B^* are used both to offset the domestic consumption distortion effect and to improve the terms of trade. In the absence of terms of trade considerations, s_B^* and s_C^* are unambiguously positive. Moreover, both s^* and t^* are positive under Bertrand or Cournot competition if demand is not too nonlinear.

In the case of homogeneous goods, $t_C = (c - C)/2 - Q_C^2 D''$, where D is the common inverse demand function. If D is linear, then $t_C^* = (c - C)/2$, the result derived in Dixit, 1984.

The case of $Q = 0$ and $q > 0$ is straightforward. The optimal policy is to provide a production subsidy until $p = c$. No tariffs are necessary to block imports. If $q = 0$ and $Q > 0$, then t_C^* and t_B^* are given by (A2) and (A4), respectively, $s_B^* = c - p(Q_B)$, and $s_C^* \leqq c - p(Q_C)$.

REFERENCES

Aczél, János, "Solution of a System of Functional-Differential Equations Arising from an Optimization Problem," *Linear Algebra and Its Applications*, November 1985, *71*, 9–13.

Boyer, Marcel and Moreaux, Michel, "Conjectures, Rationality, and Duopoly Theory," *International Journal of Industrial Organization*, March 1983, *1*, 23–41.

Brander, James A. and Spencer, Barbara J., "Tariff Protection and Imperfect Competition," in Henryk Kierzskowski, ed., *Monopolistic Competition and International Trade*, Oxford: Oxford University Press, 1984.

______ **and** ______, "Export Subsidies and International Market Share Rivalry," *Journal of International Economics*, February 1985, *18*, 83–100.

Bresnahan, Timothy F., "Duopoly Models with Consistent Conjectures," *American Economic Review*, December 1981, *71*, 934–45.

Bulow, Jeremy I., Geanakoplos, John D. and Klemperer, Paul D., "Multimarket Oligopoly: Strategic Substitutes and Complements," *Journal of Political Economy*, June 1985, *93*, 488–511.

Cheng, Leonard, "Protection of "High Technology" Industries, Countervailing Subsidies, and Welfare," mimeo., Department of Economics, University of Florida, Gainesville, 1984.

______ "Assisting Domestic Industries under International Oligopoly: The Relevance of the Nature of Competition to Optimal Policies," mimeo., Department of Economics, University of Florida, Gainesville, 1987.

Dixit, Avinash K., "International Trade Policies for Oligopolistic Industries," *Economic Journal*, Supplement 1984, *94*, 1–16.

______ "Anti-dumping and Countervailing Duties under Oligopoly," *European Economic Review*, January 1988, *32*, 55–68.

Eaton, Jonathan and Grossman, Gene M., "Optimal Trade and Industrial Policy under Oligopoly," *Quarterly Journal of Economics*, May 1986, *51*, 383–406.

Fisher, Eric O'N. and Wilson, Charles A., "International Duopoly with Tariffs," International Finance Discussion Papers No. 308, Board of Governors of the Federal Reserve System, Washington: USGPO, 1987.

Fung, K. C., "Strategic Trade Policies, Differentiated Duopoly and Intra-Industry Trade," mimeo., Economics Department, Mount Holyoke College, South Hadley, MA, 1987.

Krishna, Kala, "Trade Restrictions as Facilitating Practices," NBER Working Paper No. 1546, Cambridge, MA, NBER, January 1985.

______ **and Itoh, Motoshige,** "Content Protection and Oligopolistic Interactions," *Review of Economic Studies*, January 1988, *55*, 107–25.

Krugman, Paul, "Import Protection as Export Promotion International Competition in the Presence of Oligopoly and Economies of Scale," in Henryk Kierzskowski, ed., *Monopolistic Competition and International Trade*, Oxford: Oxford University Press, 1984.

Perry, Martin K., "Oligopoly and Consistent Conjectural Variations," *Bell Journal of Economics*, Spring 1982, *13*, 197–205.

Spencer, Barbara J. and Brander, James A., "International *R&D* Rivalry and Industrial Strategy," *Review of Economic Studies*, October 1983, *50*, 707–22.

The Missing Link: The Ricardian Approach to the Factor Endowments Theory of Trade

By ROY J. RUFFIN*

This paper views all interpersonal exchange as Ricardian, while all international trade reflects each country's factor endowment. The model provides the logical link between Heckscher-Ohlin and Ricardo. The strength of the model is that it allows simpler and more robust theorems about trade, welfare, and factor payments. For example, factor price equalization is universal and the old Mill-Ricardo theorem that the small country gains all from trade is incorrect.

David Ricardo, 1817, supposed that Portugal had a comparative advantage in wine and England a comparative advantage in cloth.[1] To generate trade, technologies had to differ between the two countries. But a factor endowment view can be achieved by assuming that the world is populated by those who have a comparative advantage in wine and those who have a comparative advantage in cloth. Let some wine makers reside in "England" and some cloth makers reside in "Portugal." Assume that those with a comparative advantage in wine or cloth are identical across countries. Without impediments to trade, there is really just the interpersonal trade between wine exporters and cloth exporters. There will be international trade provided only that the new "England" and "Portugal" claim relatively different endowments of wine exporters and cloth exporters.

In this reinterpretation of Ricardo, a technological comparative advantage is a property of factors of production and people rather than countries. The supply-side comparative advantage for a country depends on its endowment of factors of production, just as in the Heckscher-Ohlin (H-O) model. The model generates all of the main features of the H-O model of international trade: 1) trade reflects factor endowments, 2) there will be a tendency toward factor price equalization, and 3) trade benefits the abundant factor and hurts the scarce factor.

It is convenient to call the above model, "the $r \times n$ Ricardian model," where r is the number of factors and n is the number of goods. The model may be viewed in two ways. First, as indicated, it is simply the normal Ricardian model applied to interpersonal trade with international trade being the consequence of interpersonal trade across national borders. Second, the $r \times n$ Ricardian model may be regarded as a special case of the H-O model, where the inputs are perfect substitutes in the production of any good. Either view can be used as a source of economic intuition about the working of the model.

Such a model may be regarded as the "missing link" between the Ricardian and the H-O models of trade. Just as our model is a special case of the H-O model, so is the

---

*Department of Economics, University of Houston, Houston, TX 77004. This paper is dedicated to Professor Joel Sailors, who retired from the University of Houston in 1987 after more than 30 years of teaching and scholarship. I am in his debt for many discussions of the contents of this paper. Ron Jones provided detailed and insightful comments on an early draft of this work. This paper was presented at the University of California at the Santa Cruz Conference on International Economics and Finance, October 30, 1987. I have benefited from the comments of seminar participants at the University of Houston, Rice University, and Rochester University. I also wish to thank Rob Feenstra, Farhad Rassekh, Irwin Collier, and the referees for their helpful comments. I claim all residual errors.

[1] The Ricardian model has been extended to the many-country, many-good world by Frank Graham, 1948; Lionel McKenzie, 1954; and Ronald Jones, 1961, 1985. Rudi Dornbusch, Stan Fischer, and Paul Samuelson, 1977, have developed a two-country model with a continuum of goods.

traditional Ricardian model a special case of our model where each country has a monopoly on each labor type.

The Ricardian approach to the factor endowments theory of trade has the advantage of great simplicity. In contrast, the H-O theory is extremely complicated. For example: numerical solutions to even the two-by-two H-O model are beyond casual calculations; the factor price equalization theorem depends on delicate properties of the production functions and the closeness of national factor endowments; and when the number of factors exceeds the number of goods, the major theorems such as factor price equalization, Stolper-Samuelson, Heckscher-Ohlin, and Rybczynski break down or become far weaker. Even in the case of three factors and two goods, the relationship between factor endowments and trade is too complex for any simple conclusions (see Edward Leamer, 1984, pp. 16–18).[2]

Section I introduces the $r \times n$ model, describes the production transformation surface of each country or the world, and establishes the universal factor price equalization theorem. Section II illustrates the techniques of analysis with the simple 2×2 version. It formally establishes the basic theorem about trade patterns. Section III shows that trade benefits the abundant factor and hurts the scarce factor in the general case. Section IV illustrates the power of the model to shed light on old issues. I show that the John Stuart Mill theorem, John Chipman, 1965, that the small country gains the most from international trade, is specific to the Ricardian special case. In the 2×2 Ricardian model, the Mill theorem is a special case of a more general result in which the large country under certain circumstances can obtain all the gains from trade and the small country nothing. In other words, the "small country" assumption of trade theory can be misleading. Finally, Section V completely characterizes the impact of factor endowments on trade patterns when there are three factors and two goods by proving three theorems relating factor endowment to international trade. I show the exact conditions under which a "Leontief paradox" can emerge. Interestingly enough, much hinges on the "limbo" ratios discussed by Frank Graham many years ago in 1948.

I. The Model

This section describes the general characteristics of the $r \times n$ Ricardian model and compares them with the properties of the H-O model. The production transformation surface for the case of two factors and three goods is described and universal factor price equalization is demonstrated.

ASSUMPTION: *Suppose there are r types of productive factors and n goods.* We refer to each productive factor as "labor," but other interpretations are possible. The home country is endowed with L_i units of type i labor. The foreign country has L_i^* units of type i labor (variables with asterisks refer to the foreign country when possibly different). The total amount of type i labor in the world is denoted by $T_i(=L_i+L_i^*)$. Let a_{ij} denote the amount of type i labor that will produce one unit of good j. I assume the a_{ij}'s are identical all over the world. Labor supplies are internationally immobile, but nationally mobile. The amount of good j produced by L_{ij} units of type i labor is

$$X_{ij} = L_{ij}/a_{ij}. \tag{1}$$

Since good j can be produced by many labor types, total good j production would be

$$X_j = \Sigma_i L_{ij}/a_{ij}, \tag{2}$$

which shows that the Ricardian $r \times n$ model is just a special case of the H-O model with linear iso-quants. For example, with two factors and two goods the efficient allocation in an Edgeworth box would be a contract curve lying along adjacent sides from origin to origin. Since the a_{ij}'s are identical around the world, by using world factor endowments a similar Edgeworth box would describe the efficient allocation of resources for the world economy.

---

[2] There has been some work on the three-factor, two-good model. See Raveendra Batra and F. Casas, 1976; my article, 1981; Ronald Jones and Stephen Easton 1983; Katsuhiko Suzuki, 1983; and Henry Thompson, 1985. The clearest result is that the middle factor benefits from an increase in the supply of an extreme factor and the extreme factors are enemies (see my paper, 1981).

The Production Set for Each Labor Type. It is more convenient to think about the Ricardian interpretation. In Ricardo, we just consider the efficient allocation of each type of labor (or "country" in the traditional interpretation). Thus, the production set generated by the endowment and technology of each labor type is

$$(3) \quad Y_i = \{(X_{ij}) : \Sigma_j a_{ij} X_{ij} \le L_i\}.$$

Since technology is identical worldwide, a similar set defines the production set generated by the world supply of each type of labor. If Z_{ij} denotes the world output of good j by type i labor, the world production set for each type of labor is

$$(4) \quad Y_i + Y_i^* = \{(Z_{ij}) : \Sigma_j a_{ij} Z_{ij} \le T_i\}.$$

It is clear that the production transformation surface for each *labor type* is just a plane in many dimensions.

The Production Transformation Surface for the Economy. In the $r \times n$ Ricardian model each country's transformation surface consists of points (where each type of labor is completely specialized), lines (where one type of labor produces two commodities, and larger dimensional planes (where each type of labor produces more than two commodities). It is of course exactly like the world transformation surface in the multi-country version of the Ricardian theory developed by Frank Graham, 1948; Lionel McKenzie, 1954; and Ronald Jones, 1961, 1985.

Let us consider the transformation surface for a country (or the world) for the 2×3 Ricardian model. Let the three goods be 1, 2, and 3; and call the labor types or factors A and B. Figure 1 illustrates a McKenzie triangle, which gives a two-dimensional representation of the three-dimensional transformation surface. The points labeled (A_j, B_k) indicate that type A labor is specialized in good j and type B labor is specialized in good k. At the corners (A_j, B_j) for $j = 1, 2,$ and 3, the country is specialized in goods 1, 2, and 3. To determine the remaining points, we need to know the relative productivities of labor types A and B. Suppose that

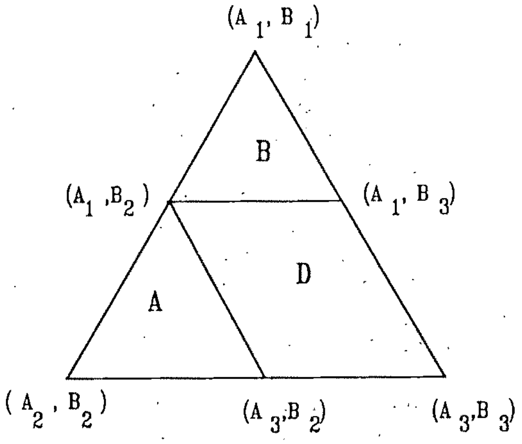


FIGURE 1

$$(5) \quad a_{A1}/a_{B1} < a_{A3}/a_{B3} < a_{A2}/a_{B2},$$

where a_{Aj}/a_{Bj} is the ratio of B's productivity in good j to A's productivity in good j. Thus, A has a comparative advantage in good 1, B in good 2, and either A or B has a comparative advantage in good 3 depending on relative wage rates. The ratio of wages, W_B/W_A, must be in between the two extremes so that factor A always exports good 1 and factor B always exports good 2. Good 3 will be produced by both factors if $W_B/W_A = a_{A3}/a_{B3}$. Factor A will produce good 3 if W_B/W_A is slightly higher than the middle productivity ratio; factor B will produce good 3 if the wage ratio is slightly lower.

On the efficient transformation surface each type of labor has its day in court (in other words, every factor has a comparative advantage). Suppose both factors are specialized in good 1 so that the production point is represented by the corner labeled (A_1, B_1). Imagine production is in the vicinity of point (A_1, B_1), where goods 2 and 3 are produced. Since type A labor has a comparative advantage in good 1, it is type B labor that first produces goods 2 or 3. As a consequence, facet B reflects the production transformation surface of type B labor. Similarly, facet A represents the production transformation surface for type A labor. It is near the (A_2, B_2) corner, because when the economy is specialized in good 2 factor A is the first to begin producing goods 1 or 3. The parallelogram labeled D is when each country is producing two goods: A labor is producing goods 1 and 3 and B labor is producing goods 2 and 3. In region D, the factor price ratio $W_B/W_A = a_{A3}/a_{B3}$.

Commodity prices, of course, determine the location of production. But production will be indeterminate except at the points. Along any line or on any facet production can be varied within prescribed limits. For instance, along the line segment from (A_1, B_1) to (A_1, B_3), factor B moves from producing good 1 to good 3. In this case, $P_1/P_3 = a_{B1}/a_{B3}$, and the price of good 2 is so low that none is produced. The line from point (A_1, B_2) to point (A_3, B_2) involves a fixed amount of good 2 production (equal to factor B's specialized output) and transferring factor A from good 1 production into good 3 production. It corresponds to a ridge on the world transformation surface. Such production indeterminacies represent no problem for the world economy, since demand conditions will pin down the exact amount of output.

Figure 1 represents the world transformation surface or the transformation surface for each country. Every country's production surface looks exactly alike except for the relative sizes of the triangles A and B and the parallelogram D. The relative sizes of triangles A and B reflect the relative supplies of A labor and B labor. In the general n-dimensional case, each country's transformation surface is the same except for the sizes of the lines, flats, and higher-dimensional subspaces.

Specialization of Goods and Factors. In our model it is possible to compare production patterns between countries that do not share the same factors. In the general H-O model, every country must have positive amounts of every factor because the production functions may not be defined with zero inputs of some factors. When inputs are perfect substitutes, new factor supplies can be added (or old ones eliminated) without affecting the efficient assignments of the other factors. For example, in Figure 1, if factor B is eliminated, then facet A just becomes the entire transformation surface.

One of the most intriguing characteristics of the $r \times n$ Ricardian model is that factors specialize more than countries. In the H-O model, if factor endowments are sufficiently different among countries, at least one of the countries must be specialized. In the $r \times n$ model, such specialization will not occur as long as a country possesses positive amounts of each factor. Clearly, the country will produce positive amounts of the goods in which its factors have a comparative advantage. This property is important because it captures one of the brute empirical facts about the world.

Universal Factor Price Equalization. The dual of the fact that the transformation surfaces between countries have flats, lines, and points that are geometrically similar is that changes in commodity prices cause parallel realignments in the activities chosen by (or assigned to) each type of labor among the various countries. Within each country, labor is allocated to its comparative advantage for any given set of world prices faced by the country. Thus, the wage rate of type i labor, W_i, is determined by the allocation that maximizes its earnings over all possible activities. If P_j is the world price of good j, it follows that

$$W_i = \max_j (P_j/a_{ij}). \tag{6}$$

THEOREM 1: *There will be universal factor price equalization as long as each country faces the same prices.*

PROOF:

From (6) it is trivial that factor prices are equalized between countries with the same P_j's and a_{ij}'s.

The factor price equalization theorem in the $r \times n$ Ricardian model releases us from such esoteric concerns as factor intensity reversals or cones of diversification. In the standard H-O model, there is factor price equalization between countries if and only if their factor endowments are in the same cone of diversification.[3] With strictly convex iso-quants, large differences in factor endowments among countries required unequal factor prices and cause different techniques of production to be adopted across countries. In the $r \times n$ Ricardian model, the cone of diversification is the entire endowment space. Thus factor endowments can differ dramatically and (as long as technologies and prices are the same the world over) factor prices remain equalized. As a consequence, there is universal factor price equalization.

Universal factor price equalization means that for any distribution of factors between countries the equilibrium will be the same as that of a fully integrated world economy, defined as the equilibrium when all goods and factors are mobile (Avinash Dixit and V. Norman, 1980). Indeed, it is obvious that as long as any factor's tastes are independent of its location, the world equilibrium vector of commodity and factor prices will be independent of the location of factors.

COROLLARY 1: *If all laborers have tastes that are independent of their location, the world equilibrium price vector is independent of the location of factors.*

This corollary follows from universal factor price equalization. To see how it works in one particular instance, suppose that the equilibrium is such that each factor is completely specialized and $r = n$. Moving any factor from one country to another causes the output of the good in which it is specialized to fall in the country of emigration and to rise in the country of immigration by exactly offsetting amounts. Thus, since the factor earns the same in both countries and world demand remains the same, the world equilibrium price vector remains unaffected. Of course, we need not assume that $r = n$ or that each factor be completely specialized. When one factor is diversified, so that a particular factor produces several goods, moving the factor from one country to the next has an indeterminate impact on the output of each country, but whatever the case the world output of each good must remain the same since world demand does not change with universal factor price equalization. For the given world demands, supply must adjust until supply equals demand at the original set of commodity and factor prices.

II. International Trade in the Two × Two Model

All exchange in the $r \times n$ Ricardian model is interpersonal. International trade is but the mere consequence of unequal distributions of the various types of labor between the home and foreign countries. The model is the purest vehicle for exhibiting the role of factor endowments in explaining trade, because all countries adopt the same technologies and the only differences observed in basic output patterns are due to factor endowment differences.

I shall now examine interpersonal and international trade in the 2×2 case. The two-by-two case allows us to introduce the basic method of analysis and can be used to illustrate the power of a simpler model of factor endowments. In this section we will just prove the H-O theorem about trade patterns with two factors and two goods.

There are two goods, goods 1 and 2, and two types of labor, type 1 and type 2. Let good 2 be the numeraire, so that p and w_i denote the price of good 1 and the wage of type i labor in terms of good 2 (that is, $p = P_1/P_2$ and $w_i = W_i/P_2$).

Interpersonal Trade. Without loss of generality assume type i labor has a comparative advantage in good i (by appropriate labeling). Thus, for example, we can speak of type 1 labor as having a comparative advantage in good 1. There will be interpersonal trade if the usual Ricardian cost ratios differ. We are supposing

---

[3] For a given set of commodity prices the cone of diversification is the set of possible factor endowments consistent with a single set of factor prices. In the two input case with two goods, the cone is formed by the tangent to the unit value iso-quants. See John Chipman, 1966.

$$a_{11}/a_{12} < a_{21}/a_{22}. \tag{7}$$

As is well known, interpersonal trade between the two labor types (before assigning the labor types to "countries") satisfies the basic laws:

$$a_{11}/a_{12} \leq p \leq a_{21}/a_{22}, \tag{8}$$

$$a_{22}/a_{12} \leq w_1/w_2 \leq a_{21}/a_{11}. \tag{9}$$

The commodity price ratio p must satisfy the range in (8) because otherwise only one good would be produced; the wage ratio w_1/w_2 must satisfy the range in (9) because otherwise one type of labor could undercut the other type in all goods.

The Heckscher-Ohlin Trade Pattern. To generate patterns of international trade it is only necessary to assume divergent factor endowments. I assume that the home country has relatively more type 1 labor. Thus,

$$L_1/L_2 > T_1/T_2 > L_1^*/L_2^*. \tag{10}$$

Obviously, the world endowment ratio T_1/T_2 must be strictly in between the individual country's endowment ratios.

Assume laborers of the same type have the same preferences all over the world. I do not assume homothetic tastes in our basic proofs. In this regard we are following my 1977 paper in which I pointed out that the assumption of homothetic tastes need not be made in the H-O theorem.[4] But in the geometric illustrations, I will make the stronger assumption that tastes are identical and homothetic between countries. While it is possible to use the former assumption in our diagrams, the additional realism is not worth the added complexity.

Suppose that initially $L_1/L_2 = T_1/T_2 = L_1^*/L_2^*$. Each country has identical production transformation curves (except possibly for scale). Since laborers of the same type have identical tastes, the aggregate tastes of each country are identical. Thus, there will then be no trade because both countries would consume and produce the same relative amounts. To determine the trade pattern with a skewed factor endowment, we need only move labor from one country to the other until trade is created (see my 1977 article). From Corollary 1 we know that this redistribution of labor does not affect the equilibrium p. Thus, the exercise tells us the pattern of trade with a skewed factor endowment. By (8) we know that at least one type of labor must be specialized. Without loss of generality, assume type 1 is specialized. Now, move a unit of type 1 labor to the home country. Then nothing happens to the world production or consumption of any good (Corollary 1). However, in the home country the output of good 1 increases by exactly the amount of good 1 produced by the unit of labor.

Trade is created when factor 1 moves because the production changes that accompany factor movements are more asymmetric than demand changes. A good is exported or imported depending upon whether demand increases less or more than supply. There are two cases, depending on whether type 2 labor is also specialized or potentially diversified. If type 2 is specialized, the movement of type 1 labor does not affect either country's production of good 2. Since the home country now has more of type 1 labor and no less of type 2 labor, the home country must import good 2 to satisfy the demands of the new immigrant. Accordingly, the home country will export good 1 and import good 2. If type 2 is diversified, so $p = a_{21}/a_{22}$, there is more potential production of good 2 by type 2 labor. In this case, both countries would produce on the corresponding flats of their respective transformation curves. Since production is indeterminate in each country, it is possible for there to be no trade or trade in either direction (for a similar problem, see Bhagwati, 1967). *We will assume that when trade is indeterminate, the physical volume of*

---

[4] See also A. D. Woodland, 1982, pp. 182 and 211, for notes on the literature and Jagdish Bhagwati and T. N. Srinivasan, 1983, for a textbook exposition.

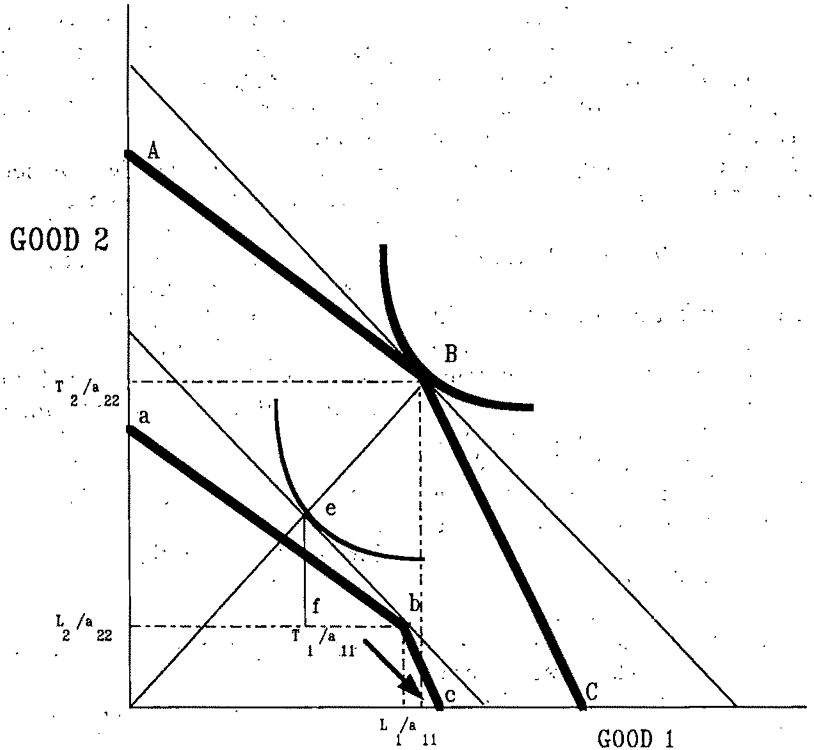


FIGURE 2

trade will be minimized. This would occur, for example, if there were an infinitesimally small tariff or trade impediment. When the unit of type 1 labor migrates from the foreign country to the home country, without any reallocations, the home country produces more good 1. There will then be an excess supply of good 1 and an excess demand for good 2. But type 2 labor is diversified: by putting more of type 2 labor into the production of good 2 its excess demand might be reduced to zero. But as the type 1 labor continues to move to the home country, type 2 labor in the home country might eventually be specialized in good 2. Any (if available) additional type 1 labor sent to the home country would then create an import demand for good 2. Thus:

THEOREM 2: *If there are two types of labor, and each type has the same tastes across countries, each country will export the good in which its abundant factor has a comparative advantage provided* (*i*) *there is international trade, and* (*ii*) *any indeterminate trade is minimized.*

Figure 2 illustrates the pattern of international trade when tastes are identical and homothetic between countries. The method is based on the idea that with identical and homothetic tastes across countries, the free trade equilibrium is represented by the tangency of a community indifference curve with the *world* transformation curve. Thus, in Figure 2, the world transformation curve is *ABC*. It is based on using world endowments and the production sets defined in (4). Because of inequality (7), the flat *AB* corresponds to type 1's relative cost and the flat *BC* reflects type 2's relative cost. Assume the free trade equilibrium is such that p is strictly in between the two cost ratios, as in Figure 2. From inequality (10), the home country's transformation curve is *abc* and has a kink at point *b* that must be to the right of *OB* since $L_1/L_2 > T_1/T_2$. The trade triangle must be *efb*, so the home country exports good 1. Note that it is only necessary to compare one country's transformation curve with the world's transformation curve.

III. Real Wages and the Terms of Trade

One of the most important propositions in international trade theory is the Stolper-Samuelson theorem that improving the terms of trade benefits the abundant factor and harms the scarce factor. This section restates the theorem for the $r \times n$ model. I also show that improving the terms of trade helps the abundant factor more than it hurts the scarce factor.

Assume that before international trade takes place the relative price of good 1 (recall $p = P_1/P_2$) will be cheaper in the home country, that is, $p < p^*$ When trade opens, p rises in the home country and p^* falls in the foreign country.

Consider what happens in the home country. Using (6) and (7), wages must be (recall $w_i = W_i/P_2$):

$$w_1 = p/a_{11} \text{ and } w_2 = 1/a_{22}. \tag{11}$$

The welfare of a type i person is given by the indirect utility function

$$u^i = v^i(p, w_i). \tag{12}$$

Let c_{ij} denote each type i person's consumption of good j. Using Roy's indentity and (11) we can see that

$$du^1/dp = v_w^1(w_1 - pc_{11})/p > 0, \tag{13}$$

as long as type 1 people consume some good 2. Using the same methods,

$$du^2/dp = -v_w^2 c_{21} < 0. \tag{14}$$

We have thus proved:

THEOREM 3: *International trade benefits the abundant factor and hurts the scarce factor.*

This theorem also shows that protection benefits the scarce factor and hurts the abundant factor, just as in the Stolper-Samuelson theorem for the standard H-O model.

In the $r \times n$ Ricardian model Theorem 3 generalizes to any number of goods and factors. In the general model, p in equation (12) would be a vector and (13) and (14) would be partial derivatives!

It is straightforward to show that the benefits of trade exceed the costs. To see this, tax type 1 labor at rate t and subsidize type 2 labor at rate s so that there is budget balance

$$tL_1 = sL_2. \tag{15}$$

Let $b = L_2/L_1$ so that $t = sb$. Thus, the indirect utilities are

$$u^1 = v^1(p, w_1 - sb), \tag{16}$$

$$u^2 = v^2(p, w_2 + s). \tag{17}$$

When trade opens, p rises in the home country. Let type 1 labor compensate type 2 labor so that $ds = c_{21}dp$. Thus, the utility of type 2 labor stays the same as they have been exactly compensated. What is the net gain to type 1 labor?

$$\begin{aligned} du^1/dp &= v_w^1[1/a_{11} - c_{11} - bc_{21}] \\ &= v_w^1 L_1[X_1 - L_1c_{11} - L_2c_{21}]. \end{aligned} \tag{18}$$

This must be positive as long as good 1 is exported! Not surprisingly, the expression is exactly the same as the expression for the change in real income when social indifference curves are used (see Richard Caves and Jones, 1981, pp. 490–91).[5]

THEOREM 4: *When the terms of trade improve, the winners can compensate the losers and the net improvement is proportional to the quantity of exports.*

---

[5] This may be compared to Peter Kenen, 1957; and Chipman and James Moore, 1972.

The above results are not surprising. However, they do illustrate the simplicity and elegance of the $r \times n$ Ricardian model.

IV. The Gains from Trade

In his seminal treatment of international values, John Stuart Mill (1929) observed:

> ...the richest countries, *ceteris paribus*, gain the least by a given amount of foreign commerce: since, having a greater demand for commodities generally, they are likely to have a greater demand for foreign commodities, and thus modify the terms of interchange to their own disadvantage.
>
> [Mill, p. 604]

I now show that this famous Mill theorem holds only in the case where each country has a monopoly on some labor type. When both countries have supplies of each factor, the country that gains the most from international trade turns out to be the one that is most different from the world economy. This is as it should be because it is well known that the gains from trade depend on the difference between autarky prices and world prices.[6] In Ricardo, the small country is less like the world economy than the large country. In my model, even with two countries, the small country may be more like the world economy. Thus, it may be the large country that trades at the small country's autarky price ratio. The Mill theorem has been so influential that the assumption that a country faces fixed international prices is called the "small country assumption." However, the results in this section show that the large country can trade at the prices of the smaller country.

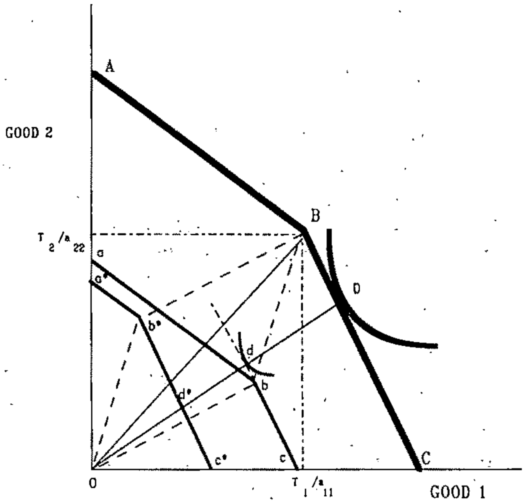


FIGURE 3

Figure 3 shows the conditions under which one country appropriates all of the gains from trade. We assume that tastes are identical and homothetic between countries, although that assumption can be relaxed. The world transformation surface is *ABC*, the home country curve is *abc*, and the foreign curve is $a^*b^*c^*$. Let *OD* be the income-consumption path for the world and also each country at the free trade equilibrium *p*. The home country will appropriate all the gains from trade if in the free trade equilibrium the home country's kink is to the right of the cone *DOB*. In particular, since the slope of *OD* is simply C_2/C_1 (where C_i is good *i* consumption), the condition is simply that

$$(19) \quad L_2a_{11}/L_1a_{22} < C_2/C_1 < T_2a_{11}/T_1a_{22}.$$

By necessity, since the world T_2/T_1 ratio is in between the country's endowment ratios, the foreign country's kink must be to the left of *OB* as long as (19) is satisfied. After trade takes place, the home and foreign countries consume at points *d* and d^*, respectively. Before trade, the home and foreign countries consume at points *b* and d^*, respectively. It is clear that point *b* is on a lower indifference curve than point *d*. In Figure 3, we have drawn the diagram so that the ratio of home country GNP to foreign GNP is $Oc/Oc^* > 1$. It is easy for the larger home country to receive all the gains from trade.

---

[6] Murray Kemp (1969, pp. 266–67) argues that this statement is valid if "factor supplies are inelastic and if for any reason the community behaves as if it were trying to maximize a single concave utility function."

Thus:

THEOREM 5: *If tastes are identical and homothetic between countries and the home country is richly endowed with type* 1 *labor* [*condition* (10) *is satisfied*], *then the home country receives all of the gains from trade if and only if condition* (19) *is satisfied in the free trade equilibrium. The size of country is irrelevant.*

In Figure 3, the free trade price ratio p equals the cost ratio for labor type 2. In effect, there is so much type 2 labor in the world that world prices conform to type 2's labor costs. Thus, type 2 labor gains nothing from interpersonal exchange and type 1 labor gains all. But in Figure 3, the home country is well endowed with type 1 labor and the foreign country is well endowed with type 2 labor. Thus, if one type of labor does not gain from inter*personal* trade, then it must be that the country which is richly endowed with that type of labor will not gain from inter*national* trade.

An example will illustrate that it is the comparison of factor endowments with the world factor endowments that determines the division of the gains from trade. Following Chipman, 1965, let us use the Mill-Graham demand assumption that $u = C_1C_2$. By examining the slope of the indifference curve through point B relative to the slopes of the AB and BC flats in Figure 3, it is easy to show that the free trade price ratio will be the following step function:

$$(20)\quad p = \begin{cases} a_{11}/a_{11} & \text{if } T_2/T_1 < a_{22}/a_{12}, \\ T_2a_{11}/T_1a_{22} & \text{if } a_{22}/a_{12} \leq T_2/T_1 \\ & \quad \leq a_{21}/a_{11}, \\ a_{21}/a_{22} & \text{if } T_2/T_1 > a_{21}/a_{11}. \end{cases}$$

Figure 4 shows a graph of function (20) relating the factor endowment ratio to the commodity price ratio. If a country's endowment ratio is used, the function shows the autarky price; if the world endowment ratio is used, the function shows the world free trade price.

THEOREM 6: *Under the Mill-Graham demand assumption and condition* (10), *the home country will gain all from international trade and the foreign country will gain nothing if only if*

$$(21)\quad L_2^*/L_1^* > T_2/T_1 > a_{21}/a_{11} > L_2/L_1.$$

As long as (21) *holds, the size of country is irrelevant.*

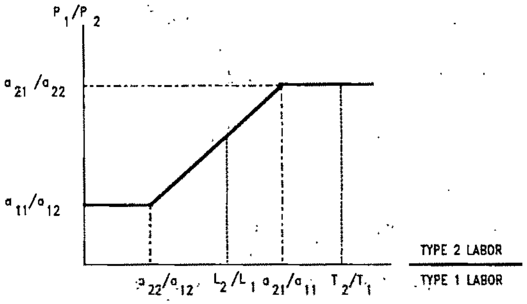


FIGURE 4

When (21) holds, before trade the home country's p will be less than a_{21}/a_{11} (see Figure 4) and the foreign country's autarkic p^* will equal a_{21}/a_{11}. However, in the free trade equilibrium p equals the foreign country's autarkic p^*.

A numerical example vividly illustrates how large the home country can be and still capture all the gains from trade. Suppose $a_{11} = a_{22} = 1$ and $a_{21} = a_{12} = 2$. Assume $L_2 = 99$, $L_1 = 50$, $L_2^* = 4$, and $L_1^* = 1$. Condition (21) is satisfied so the foreign country gains nothing from trade. However, at the free trade price ratio home GNP is 199 and foreign GNP is only 6 (from (20) $p = 2$ and from (11) $w_1 = 2$ and $w_2 = 1$). Thus, the home country is over 33 times as large as the foreign country and yet receives all the benefits of trade!

This insight is important from an empirical standpoint. Trade theorists often assume that just because trade is more important for small countries than large countries that the former gains more than the latter. However, Theorems 5 and 6 imply that the proportion of exports to GNP is not what matters.

To show that Theorem 6 is more fundamental than Mill's theorem, it is instructive to derive Mill's theorem from condition (21).

MILL'S THEOREM: *If the utility functions are Mill-Graham, the home country has a monopoly on type* 1 *labor, and the foreign country a monopoly on type* 2 *labor, then the home country receives all the gains from trade if and only if it has a smaller GNP at world prices.*

To establish this, if $L_2 = 0$ and $L_1^* = 0$, condition (21) becomes

$$T_2/T_1 = L_2^*/L_2 > a_{21}/a_{11}. \tag{21'}$$

By (20) $p = a_{21}/a_{22}$. Thus, by (11), $w_1 = a_{21}/a_{22}a_{11}$ and $w_2 = 1/a_{22}$. GNP in the home country is $y = (a_{21}/a_{22}a_{11})L_1$ and GNP in the foreign country is $y^* = L_2^*/a_{22}$. The home country gains all and the foreign country gains nothing if and only if condition (21′) holds. But $y < y^*$ if and only if $L_2^*/L_1 > a_{21}/a_{11}$ holds.

V. The Three-by-Two Ricardian Model

The simplicity of the above model allows us to explore the role of a third factor. It has long been argued (Jaroslav Vanek, 1963) that the existence of natural resources along with capital and labor helps explain the evolution of U.S. foreign trade. In the specific-factors model, it is possible to trace the impact of factor endowment on trade patterns (see Jones, 1971, and Paul Samuelson, 1971). But the impact of factor endowments on trade is very difficult in the traditional H-O framework.

Let there be three labor types: type 1, type 0, and type 2. By suitable numbering we can always write

$$a_{11}/a_{12} < a_{01}/a_{02} < a_{21}/a_{22}. \tag{22}$$

Thus, type 1 always has a comparative advantage in good 1 and type 2 always has a comparative advantage in good 2. Type 0 labor is the "middle" factor in my 1981 paper. The comparative advantage of type 0 labor depends on whether p exceeds, equals, or falls short of the middle cost ratio. I will call a_{01}/a_{02} the middle ratio. I shall also call type 0 labor the middle factor.

The price ratio p in equilibrium must be in between the extreme cost ratios. When p does not equal any cost ratio, Graham aptly described the equilibrium as being in "limbo" (see Graham, 1948, p. 52). Thus, we define a *limbo ratio as any* p *not equal to a cost ratio.*

How does the third factor shape the patterns of international trade? Without loss of generality, we assume

$$L_1/L_1^* > L_2/L_2^*. \tag{23}$$

The home country is still relatively well endowed with type 1 labor compared to type 2 labor. With type 0 labor, we have the three alternatives:

$$L_1/L_1^* > L_0/L_0^* > L_2/L_2^*, \tag{24}$$

$$L_0/L_0^* > L_1/L_1^* > L_2/L_2^*, \tag{25}$$

$$L_1/L_1^* > L_2/L_2^* > L_0/L_0^*. \tag{26}$$

It will be useful to refer to a factor whose factor supply ratio is in between the extremes as one that is "widely distributed."

THEOREM 7: *Assume conditions* (22) *and* (23) *hold. If laborers of the same type have the same taste, indeterminate trade is minimized, there is actual international trade, and condition* (24) *is satisfied, then the home country will export good* 1 *and the foreign country will export good* 2.

Start with the no-trade equilibrium, where $L_1/L_1^* = L_0/L_0^* = L_2/L_2^*$. By Corollary 1, we know that moving factor endowments from one country to the next does not affect the equilibrium p. To satisfy condition (24), where the middle factor is widely distributed, we merely move a bit of type 1 labor to the home country and a bit of type 2 labor to the foreign country. As in Theorem 2, the key idea in the proof is that when factors move output changes are more asymmetric than demand changes. Consider first the case in which p is in limbo. In that case, both labor types will be specialized. Thus, in the home country, the output of good 1 rises and the output of good 2 falls. The opposite happens in the foreign country. Accordingly, the home country must now export good 1 and the foreign country must export good 2. Consider now the case where p is not in limbo. If p equals one of the cost ratios, then one labor type will be diversified. Without loss of generality, suppose $p = a_{21}/a_{22}$, so that type 2 labor is diversified. Moving type 1 labor to the home country need not create trade because of compensating shifts in the production of good 1 by type 2 labor. But if enough type 1 labor is sent to the home country until type 2 labor is specialized in good 2 in the home country, as in Theorem 2, then if minimal trade exists it must be that the home country exports good 1.

Figure 5 illustrates Theorem 7 for the case in which p is in limbo and tastes are identical and homothetic between countries. The world transformation curve is $ABCD$; the home country transformation curve is $abcd$. The home country is abundant in factor 1 relative to factor 2 because the line segment ab is proportionately longer than AB and the line segment cd is proportionately shorter than CD. As long as condition (24) is satisfied, the home country's production point c must be southeast of its consumption point e.

Now let us consider the case in which the world equilibrium p equals the middle factor's cost ratio.

THEOREM 8: *Assume conditions* (22) *and* (23) *hold. If laborers of the same type have the same taste, indeterminate trade is minimized, there is actual international trade, and* $p = a_{01}/a_{02}$, *then the home country will export good* 1 *and the foreign country will export good* 2.

To prove this we can apply Theorem 7. Suppose we start with condition (24). We thus obtain the trade pattern that the home country exports good 1 provided there is minimal trade. It is possible, of course, for the equilibrium to involve no trade. But the statement of the theorem implicitly rules out this case by having factor endowments far enough apart. By redistributing the middle factor (type 0 labor) between the countries we can generate conditions (25) or (26) from the initial position. The intuition of the result is clear: redistributing type 0 labor has no impact on the trade pattern if $p = a_{01}/a_{02}$ because type 0 labor is then self-sufficient. Being self-sufficient, type 0 labor has no impact on international markets. Thus, no matter what the supply or distribution of type 0 labor, the trade pattern will be unaltered by moving from (24) to (25) to (26).

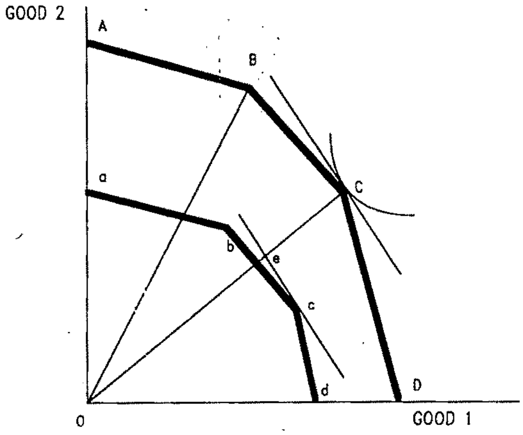


FIGURE 5

I now turn to a result intermediate between Theorems 7 and 8. This result gives us a necessary condition for a Leontief paradox. In this context, by a Leontief paradox I mean that the home country is relatively abundant in type 1 labor relative to type 2 labor (i.e., condition (23) holds), but that the home country nevertheless imports good 1.

THEOREM 9: *Assume conditions* (22) *and* (23) *hold. If laborers of the same type have identical tastes and indeterminate trade is minimized, then the home country exports good* 1 *and the foreign country exports good* 2 *if* (*i*) $p < a_{01}/a_{02}$ *and conditions* (24) *or* (26) *hold; or* (*ii*) $p > a_{01}/a_{02}$ *and conditions* (24) *or* (25) *hold.*

We only prove part (*i*)–part (*ii*) is symmetrical. Start with (24) so that from Theorem 7 we know the foreign country (which is

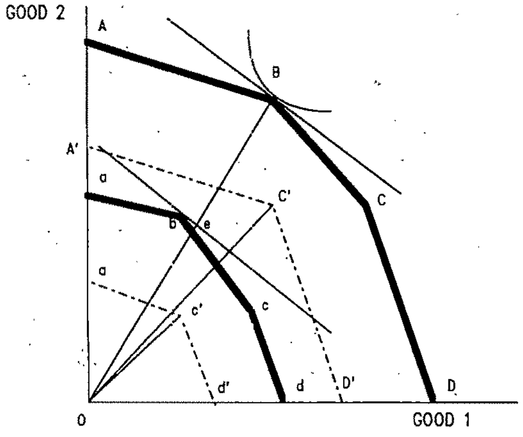


FIGURE 6

relatively abundant in type 2 labor compared to type 1 labor) exports good 2. Since the value of good 1 is less than type 0's relative cost, the middle factor has a comparative advantage in good 2. Now redistribute the middle factor to the foreign country until condition (26) becomes relevant. Since the foreign country is receiving more of a factor that exports good 2, its exports of good 2 must increase. Thus, the trade pattern will remain the same.

Trade pattern reversals are clearly possible when p does not equal the middle cost ratio and enough of the middle factor is available to reverse the trade pattern. In the case of $p < a_{01}/a_{02}$, by sending the middle factor to the home country, the trade pattern can be reversed.

In Figure 6, an example of a Leontief paradox is illustrated. The world transformation curve is *ABCD*; the home transformation curve is *abcd*. The ray *OB* is the common income-consumption path. Clearly, the home country imports good 1. But by eliminating the middle factor, we can see that $L_1/L_2 > T_1T_2 > L_1^*/L_2^*$. The dotted transformation curves *A'C'D'* and *a'c'd'* show the world and home transformation curves without the middle factor. Since *OC'* is steeper than *Oc'*, it must be that without the middle factor the home country would export good 1.

VI. Conclusions

I have shown that the $r \times n$ Ricardian model generates all of the essential features of H-O: trade reflects factor endowments, trade brings about factor price equalization, and trade benefits the abundant factor and hurts the scarce factor. The advantages of the $r \times n$ Ricardian model are many. First, it offers a stripped-down version of the factor endowments theory that can be used for introducing complications such as uncertainty or product differentiation. For example, on my own research agenda I plan to examine the consequences of costly information regarding the comparative advantages of different factors. Second, unlike H-O, numerical solutions are easily constructed. This should make an algorithm for computing world equilibrium in the $r \times n$ case quite feasible. Third, the model can be used to discuss the determinants of trade in a three-factor context—a difficult problem in the H-O model. Fourth, the model generates clear-cut relationships between commodity prices and real factor rewards, unlike the H-O model when the number of factors exceeds two. Fifth, the model suggests new ways of looking at the gains from trade in relation to the size of country. Sixth, it is likely that the simpler approach to the factor endowments theory might suggest theorems that may hold in the more general context of an H-O model.

REFERENCES

Batra, Raveendra and Casas, F., "A Synthesis of the Heckscher-Ohlin and the Neoclassical Models of International Trade," *Journal of International Economics*, February 1976, *6*, 21–38.

Bhagwati, Jagdish, "The Proofs of the Theorems on Comparative Advantage," *Economic Journal*, March 1967, *77*, 75–83.

______ **and T. N. Srinivasan,** *Lectures on International Trade*, Cambridge: MIT Press, 1983.

Caves, Richard and Jones, Ronald W., *World Trade and Payments*, 3rd ed., Boston: Little, Brown, 1981.

Chipman, John S., "A Survey of the Theory of International Trade: Pt. I, The Classical Theory," *Econometrica*, July 1965, *33*, 477–519.

______, "A Survey of the Theory of International Trade: Pt. III, The Modern Theory," *Econometrica*, January 1966, *34*, 18–76.

______ **and Moore, James C.,** "Social Utility and the Gains from Trade," *Journal of International Economics*, May 1972, *2*, 157–72.

Dixit, Avinash K. and Norman, V., *Theory of International Trade*, London: Cambridge University Press, 1980.

Dornbusch, Rudi, Fischer, Stanley and Samuelson, Paul, "Comparative Advantage, Trade, and Payments in a Ricardian Model with a Continuum of Goods," *American Economic Review*, December 1977, *67*, 832–39.

Graham, Frank D., *The Theory of International Values*, Princeton: Princeton University Press, 1948.

Heckscher, Eli, "The Effect of Foreign Trade on the Distribution of Income," *Ekonomisk Tidskrift*, 1919, 497–512; reprinted as Chapter 13 in *A. E. A. Readings in the Theory of International Trade*, Philadelphia: Blakiston, 1949.

Jones, Ronald W., "Comparative Advantage and the Theory of Tariffs: A Multi-Country, Multi-Commodity Model," *Review of Economic Studies*, June 1961, *28*, 161–75.

______, "A Three-Factor Model in Theory, Trade, and History," in J. N. Bhagwati et al., eds., *Trade, Balance of Payments, and Growth: Essays in Honor of Charles P. Kindleberger*, Amsterdam: North-Holland, 1971.

______ **and Easton, Stephen T.,** "Factor Intensities and Factor Substitution in General Equilibrium," *Journal of International Economics*, August 1983, *15*, 65–99.

______, "Ricardo-Graham World Transformation Surfaces," *Osaka City University Economic Review*, Spring 1985, 27–33.

Kemp, Murray, *The Pure Theory of International Trade and Investment*, Englewood Cliffs, NJ: Prentice-Hall, 1969.

Kenen, Peter B., "On the Geometry of Welfare Economics," *Quarterly Journal of Economics*, August 1957, *71*, 426–77.

Leamer, Edward E., *Sources of Comparative Advantage: Theory and Evidence*, Cambridge: MIT Press, 1984.

McKenzie, Lionel W., "Specialization and Efficiency in World Production," *Review of Economic Studies*, June 1954, *21*, 165–80.

Ricardo, David, *On the Principles of Political Economy and Taxation*, London: John Murray, 1817. Reprinted in *The Works and Correspondence of David Ricardo*, Pierro Sraffa, ed., Vol. I, New York: Cambridge University Press, 1951.

Ruffin, Roy J., "A Note on the Heckscher-Ohlin Theorem," *Journal of International Economics*, November 1977, *7*, 403–06.

______, "Trade and Factor Movements with Three Factors and Two Goods," *Economic Letters*, 1981, *7*, 177–82.

Samuelson, Paul, "Ohlin Was Right," *Swedish Journal of Economics*, December 1971, *73*, 365–84.

Suzuki, Katsuhiko, "A Synthesis of the Heckscher-Ohlin and the Neoclassical Models of International Trade: A Comment," *Journal of International Economics*, February 1983, *14*, 141–44.

Thompson, Henry, "Complementarity in a Simple General Equilibrium Model," *Canadian Journal of Economics*, August 1985, *18*, 616–21.

Vanek, Jaroslav, *The Natural Resource Content of United States Foreign Trade, 1870–1955*, Cambridge: MIT Press, 1963.

Woodland, A. D., *International Trade and Resource Allocation*, Amsterdam: North-Holland, 1982.

Hysteresis in Import Prices: The Beachhead Effect

By RICHARD BALDWIN*

This paper shows that temporary exchange rate fluctuations can have persistent (i.e., hysteretic) effects on trade prices and quantities. Specifically, if market-entry costs are sunk, sufficiently large real exchange rate shocks can alter domestic market structure and thereby induce hysteresis. This simple idea has strong implications for exchange rate theory, trade policy, and the estimation of trade equations. Empirical evidence suggests that the recent dollar overvaluation induced hysteresis in U.S. import prices.

International economists typically assume that temporary real exchange rate shocks can have only temporary real effects—and no effect at all on the underlying structure of the economy. For instance, in the well-known Rudiger Dornbusch (1976) overshooting model, a one-time money stock increase leads to a temporary real depreciation. However, in the long run the shock is assumed to be neutral, that is, to have no real effects. In the empirical trade literature, researchers implicitly assume that their trade volume and price equations are stable through time. In particular, they ignore the possibility that exchange rate shocks might themselves cause structural breaks in the equations.

This paper shows that, even in a simple "off-the-shelf" industrial organization model, this assumption is unfounded; if market-entry costs are sunk, exchange rate shocks can alter domestic market structure and thereby have persistent real effects. In other words large, temporary exchange rate shocks may result in hysteresis in import prices and quantities.[1] This simple idea has strong implications for exchange rate theory, for trade policy, and for the estimation of trade equations as the present paper shows.

Moreover, to show that the theoretical point is not just empirically empty theorizing, this paper presents evidence which suggests that the recent dollar overvaluation is an example of a hysteresis-inducing shock. The real U.S. dollar exchange rate (based on wholesale price data) rose about 20 percent in the 18 quarters leading up to 1985:1, and fell almost as much in the subsequent ten quarters. This dollar cycle has had rather puzzling effects on import prices. The appreciation lowered real import prices, and the depreciation has partially forced them back up. However, since the early 1980s, import prices appear to have been below the level predicted by the historical relationship. This paper formally shows that the pass-through relationship (of real exchange rates to real import prices) has indeed shifted in the 1980s, and that the nature and timing of the shift is consistent with the market structure changes predicted by the model. A direct test of the model is not performed due to data limitations.

The notion that large exchange rate changes have a qualitatively different impact

---

*Columbia University Business School, Columbia University, New York, NY 10027. I gratefully acknowledge the comments and suggestions of Paul Krugman, Avinash Dixit, Gene Grossman, Jeff Sachs, Rob Feenstra, Ron Jones, Alan Stockman, Steve Ambler, Cathy Mann, Harry Foster, Nils Gottfries, Assar Lindbeck, Rich Lyons, Jim Levinshon, Alberto Giovannini, and Bill Helkie.

[1] Hysteresis is the failure of a property changed by an external agent to return to its original value when the cause of the change is removed.

on trade than do small changes is not novel. Guy Orcutt (1950) conjectures that large price changes have a quantum effect on import volume. Krugman (1986) conjectures that the strong dollar-induced hysteresis in the U.S. trade balance is due to dynamic economies of scale.

A preliminary draft of this paper, Baldwin (1986), presents two versions of a model (the beachhead or sunk-cost model) which displays hysteresis in import quantities. The present paper extends the latter by allowing for an infinite horizon, a more general demand structure, and the presence of home firms. Moreover, it shows that hysteresis can occur in import prices, as well as quantities.

A number of papers have extended the original beachhead framework. Baldwin and Krugman (1986) allow for a stochastic exchange rate process, and more importantly shows that hysteresis in imports leads to hysteresis in the equilibrium exchange rate. Dixit (1987a, 1987b) uses a related setup to show that quantity hysteresis can occur when the exchange rate follows a continuous-time random walk. Harry Foster and Baldwin (1986) examine a model of marketing capacity constraints in which hysteresis in quantities occurs. Charles Bean (1987), using a modified beachhead model, finds evidence that the 1978–1981 sterling overvaluation had hysteretic effects on British exports.

The paper is organized in four sections. The first presents the model. The second studies the positive effects of exchange rate changes. The third presents some empirical evidence. The last section presents some conclusions.

I. The Beachhead Model

The basic economics of the model consists of two assertions: (*i*) real exchange rate shocks can alter a country's market structure, and (*ii*) market structure affects import prices and volumes. We use a modified Michael Spence (1976) and Dixit and Joseph Stiglitz (1977) (S-D-S) framework to model the industry structure and imperfect competition in a partial equilibrium setting. Foreign and home firms engage in Cournot competition in the domestic market for a particular good with each firm selling a different S-D-S variety.

Home and foreign production costs (c measured in home currency and c^* measured in foreign currency) are linear homogeneous in output. Firms must also incur a fixed, sunk market-entry cost, F, which reflects the cost of the firm-specific and market-specific assets that are required to sell in the market. For example, F could represent the costs of setting up a distribution and service network, of establishing a brand name through advertising, or of bringing the foreign product into conformity with domestic health and safety regulations. The results would go through as long as at least part of F is sunk.

Each period the firm is active, the sunk investment requires maintenance represented by a fixed maintenance cost, G. For example if F represents the cost of a brand name introduction advertising blitz, G would represent the brand name maintenance advertising. F and G are incurred in the home country and so are independent of the exchange rate.[2] If G is not spent the sunk asset disappears, that is, the firm exists.[3] We assume that $F > G$.

Since firms are making intertemporal decisions, the exact nature of firm's exchange rate expectations are crucial to the mechanics of the model. The most complete approach would be to specify a macro model (endogenizing the real exchange rate) with which firms would form their expectations. In many such macro models, persistent changes in domestic (or foreign) monetary fiscal policies can result in perfectly anticipated real exchange rate shocks.[4] This paper omits the macro model, simply assuming that firms perfectly anticipate the exchange rate path. This assumption is not crucial to the hysteresis result.[5]

---

[2] If F and G involved foreign currency costs, appreciations would lower entry costs as well as marginal costs, making entry more likely.

[3] Allowing for the possibility that not all the sunk asset disappears when the firm exits, would imply that the reentry costs are below entry costs. This consideration would complicate the analysis without providing any compensating insight.

[4] For example, Dornbusch (1976).

We limit our attention to shocks where the exchange rate equals e^0 in period zero, jumps in period 1 to E, remains there for $T-1$ periods, and returns to e^0 in period T for all future periods. We refer to this stylized shock as $V(E,T)$. An overvaluation is denoted as $V(A,T)$, and an undervaluation as $V(D,T)$.

Edward Chamberlain (1933) argues that an increase in the number of varieties in an industry shifts down and makes more elastic the demand curve for each variety. We therefore include the total number of S-D-S varieties sold by home and foreign firms in the domestic market, m_t, as an argument in the inverse demand functions. In the spirit of the S-D-S model, the inverse demand function for each firm is identical although each is for a different variety.

Formally a typical foreign firm chooses sales, y_t (all t), to maximize:

$$\sum_{t=0}^{\infty} R^t(P[m_t, y_t]y_t - c_t^* e_t y_t - G) - F,$$

where $R=1/(1+r)$, r is a constant discount rate, and e_t is the exchange rate (domestic currency per foreign currency). Since e_t is deterministic, the issue of whether the foreigners maximize profits in home or foreign currency is moot. By S-D-S symmetry, y_t is identical for all foreign firms. Home firms chose sales x_t (all t) to maximize:

$$\sum_{t=0}^{\infty} R^t(P[m_t, x_t]x_t - c_t x_t - G) - F.$$

We assume the demand curve is such that marginal revenue is decreasing in sales so that the second-order condition is met. Clearly the home-operating profit function (the profit function exclusive of F). OΠ, has m_t as its argument. The foreign operating profit function, $O\Pi^*[e_t, m_t]$, has m_t and e_t as arguments. We assume OΠ and OΠ* are smooth and decreasing in their respective arguments.

The period zero equilibrium is taken as given and plays the role of an initial condition. A prospective entrant calculates what its discounted profits would be if it entered (choosing sales optimally, based upon a period-by-period Cournot-Nash equilibrium) allowing of course for the possibility that it may exit in the future. If discounted profits are sufficient to cover F, the firm enters. Incumbent firms choose sales optimally, and if the anticipated revenues are enough to cover variable costs, they remain in the market. Since the period-by-period sequential entry equilibrium concept is Cournot-Nash, the multiperiod equilibrium is subgame perfect. Ignoring integer constraints, defining

$$S_\tau = \sum_{i=0}^{\infty} R^i(O\Pi_{\tau+i}[m_{\tau+i}]), \quad \text{and}$$

$$S_\tau^* = \sum_{i=0}^{\infty} R^i(O\Pi_{\tau+i}^*[e_{\tau+i}, m_{\tau+i}]),$$

the entry and exit conditions for home and foreign firms in period τ are, respectively,

$$(1) \qquad S_\tau > F, \quad S_\tau^* > F,$$

$$(2) \qquad S_\tau < 0, \quad S_\tau^* < 0.$$

Clearly there is a gap between the entry and exit conditions. The gap implies that there is not a unique number of firms in the period-by-period equilibria. This multiplicity of equilibria is the key to hysteresis. As we shall see below, a large enough $V(E,T)$ will permanently change the market structure, resulting in permanent real effects.

II. The Equilibrium

Before studying the effects of exchange rate shocks, we calculate the equilibrium for a constant, or benchmark, exchange rate, e^0. For convenience, we assume that e^0 is such that $c = e^0 c^*$.[6] For this e^0, S_τ and S_τ^* equals S_b which is in turn a function only of a

---

[5] Baldwin and Krugman (1986), and Dixit (1987a) make alternative assumptions.

[6] This assumption reduces the number of cases which must be addressed in determining the impact of $V(E,T)$ on m, but it does not alter the main results.

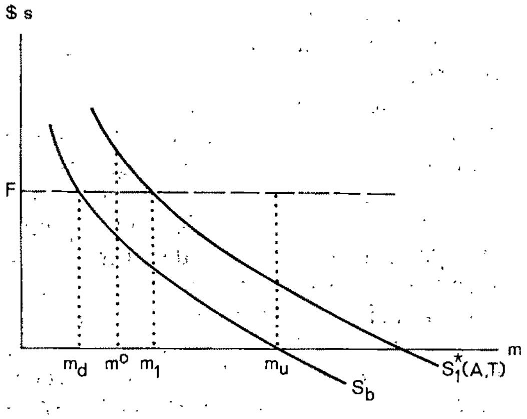


FIGURE 1. EQUILIBRIUM RANGE OF FIRMS

constant m, m_b (m is constant since e is). Figure 1 facilitates the determination of m_b. $S_b[m]$ is decreasing in m since OΠ and OΠ* are. For $m < m_d$, (1) holds so firms would enter. For $m > m_u$, (2) holds so some firms would exit. Any m between m_u and m_d constitutes an equilibrium since neither (1) nor (2) is binding.

Foreign and home firms' sales, y_b and x_b, respectively, are the solutions to the Euler equations

$$
\begin{aligned}
(3)\quad 0 &= P[m_b, y_t] \\
&\quad + y_t \partial P[m_b, y_t]/\partial y_t - c_t^* e_t, \\
0 &= P[m_b, x_t] \\
&\quad + x_t \partial P[m_b, x_t]/\partial x_t - c_t,
\end{aligned}
$$

The benchmark prices, P_b^* and P_b are equal to $P[m_b, y_b]$ and $P[m_b, x_b]$, respectively.

A. *Competitiveness Effects of Exchange Rate Shocks*

Next we study the link between exchange rate fluctuations and m. As we shall show, $V(E,T)$ fall into two types: $V(E,T)$ which cause no entry or exit (we define these as small exchange rate changes), and those which do (we call these large exchange rate changes). Large $V(E,T)$ themselves fall into two categories: those where m^0 changes in period 1 and stays permanently at the new level, and those where m^0 jumps up in period 1 and then falls back to m_u.

To facilitate the characterization of the exchange shock-market structure link, we define S_τ^* during an undervaluation and overvaluation as $S_\tau^*(D,T)$ and $S_\tau^*(A,T)$ respectively. Several properties of $S_\tau^*(A,T)$ and $S_\tau^*(D,T)$ are of interest: (i) they are monotonically decreasing in A and D, respectively (for all $\tau < T$); (ii) $S_\tau^*(A,T)$ is increasing in T, and $S_\tau^*(D,T)$ is decreasing in T; and (iii) $S_\tau^*(A,T)$ is nonincreasing, and $S_\tau^*(D,T)$ is nondecreasing through time (for all $\tau > 1$). Properties (i) and (ii) follow directly from the definition of $V(E,T)$ and the properties of OΠ*(e_t; m_t). To demonstrate (iii), suppose on the contrary, $S_{T-i-1}^*(A,T) < S_{T-i}^*(A,T)$. Since $e_{T-i-1} \leq e_{T-i}$, the supposition can hold only if $m_{T-i} < m_{T-i-1}$. This in turn implies exit occurs in period $T-i$, so $S_{T-i}^*(A,T) = 0$, or $S_{T-i} = 0$. But then the supposition (or $m_{T-i} < m_{T-i-1}$) implies that $S_{T-i-1}^*(A,T)$ (or S_{T-i-1}) is negative. This is a contradiction, since exit ensures that they remain nonnegative. A symmetric argument shows that $S_\tau^*(D,T)$ is nondecreasing with time.

We consider first the set of small $V(E,T)$. We can ignore S_τ, since $V(E,T)$ affects it only through changes in the m_t's. The set's end points, A' and D', are given by $S_1^*(A',T) = F$, and $S_1^*(D',T) = 0$ for $m_1 = m^0$. A' and D' depend on m^0 as well as on T. Specifically by (i) and (ii), the longer the overvaluation is, the closer are A' and D' to e^0, and the closer m^0 is to $m_u(m_d)$, the closer is $A'(D')$ to e^0.

Property (i) implies that for all $V(E,T)$, where $A' < E < D'$, neither the entry nor exit condition are binding in period 1. For any small $A < e^0$, $S_\tau^*(A,T)$ falls over time up to $\tau = T$ where $S_T^*(A,T) = S_b(m^0)$. For small $D > e^0$, $S_\tau^*(D,T)$ rises over time to $S_b(m^0)$. Clearly then neither (1) nor (2) ever binds, so m will not change. This finishes our characterization of small exchange rate shocks.

Next we characterize the first type of large $V(A,T)$. Figure 1 shows an example. S_1 depends only on m's so $S_1 = S_b$. The $V(A,T)$ shifts S_0^* up to S_1^*, thus driving m^0 to m_1. By (iii), S_τ^* shifts back to S_b, yet m remains at m_1 since neither (1) nor (2) holds after $\tau = 1$. We shall show that any $V(E,T)$ which causes m^0 to jump to an m_1 within the $m_d - m_u$ range is a member of this first type. If m is to jump to m_1 and stay there, then (1) and (2) must not bind after $\tau = 1$. By definition of a large $V(A,T)$, $S_1^*(A,T) = F$. By (*iii*), $S_\tau^*(A,T) < F$ after $\tau = 1$, so (1) does not hold regardless of the size of m_1. By definition of $V(A,T)$, $S_T^*(A,T) = S_b$. Thus if $m_1 < m_u$, $S_\tau^*(A,T) > 0$ for all τ, so (2) never binds. When $m_1 = m_u$, then $S_T^*(A,T) = 0$ with $m = m_1$. Similar reasoning indicates that for all $V(D,T)$ where $m_1 > m_d$, m jumps down to m_1 and remains there. In both cases, home firms never enter or exit since m stays in the $m_d - m_u$ range.

We shall show that the second category of large shocks involves $V(A,T)$ which cause m_1 to be greater than m_u. If $m_1 > m_u$, then m must fall to m_u by period T. Otherwise $S_T^*(A,T)$ would be negative. Property (*iii*) implies that m falls from m_1 to m_u, and then remains there. For such $V(A,T)$ all home firms exit, since with $m_1 > m_u$ $S_\tau < 0$. Note that no $V(D,T)$ can drive m_1 below m_d because of home firm entry.

B. *Price Effects of Small and Large Exchange Rate Changes*

To summarize, small $V(E,T)$ causes no change in market structure, while large $V(E,T)$ do. An exchange rate shock can be "large" when it is very big (E is very different from e^0) or very long (T is large). Large shocks can change the size and home/foreign composition of m. We turn next to the effects of exchange rate movements on the price of imports.

Rearranging (3) implies import prices are related to e by

$$P_t = \frac{1}{1 - 1/\varepsilon[m_t, y_t]} c_t^* e_t. \tag{4}$$

The perceived elasticity, ε, is a function of m (due to the Chamberlain assumption) and y (to allow for the nonconstancy of the demand elasticity).

The time paths of P, e, and m are depicted in Figure 2A for a small exchange rate shock. At time zero all firms realize that e will follow the path shown. In response they lower import prices. Since the shock is small (i.e., $A' < E < e^0$) there is no change in m so that when the exchange rate returns to e^0, P returns to its original value. Figure 2B shows the time paths for a large shock. Upon announcement of the overvaluation, m jumps up. The price falls due to the marginal cost reduction (lower e reduces foreign costs measured in home currency), and to the market structure change (more competition forces down profits margins). After the overvaluation passes, the marginal costs return to their original level. However, m is still higher so the post-shock price is permanently lower than the pre-shock price. This is hysteresis.

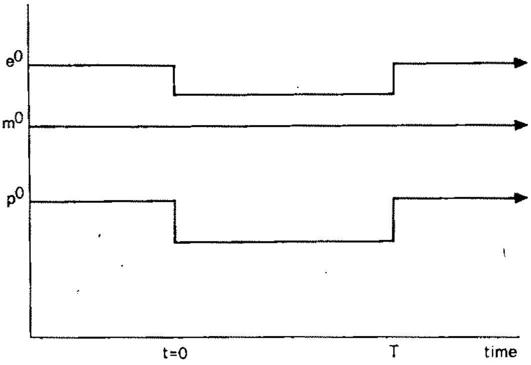


FIGURE 2A. TIME PATHS OF e, m, P FOR SMALL OVERVALUATION

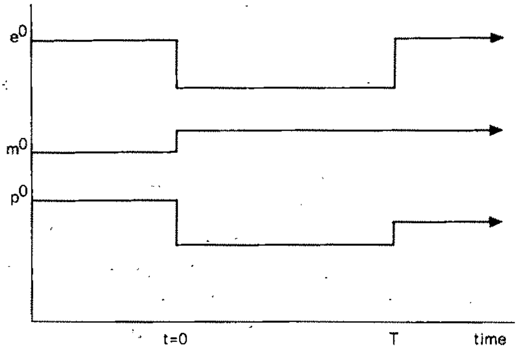


FIGURE 2B. TIME PATHS OF e, m, P FOR LARGE OVERVALUATION

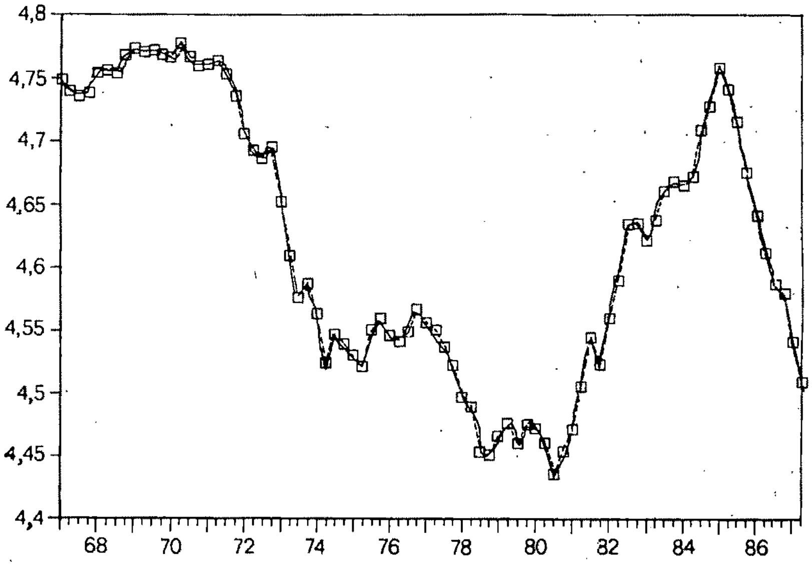


FIGURE 3. THE U.S. DOLLAR REAL EXCHANGE RATE

III. Empirical Evidence

The beachhead model argues that a large enough appreciation can induce entry and that the presence of additional entrants can affect pricing behavior. In a standard time-series regression of the relationship between the exchange rate and import prices this event would appear as a structural break. The unprecedented magnitude and duration of the 1980s' real dollar shock provides one opportunity to test the predictions of the model.

Clearly, earlier real exchange rate shocks could conceivably have induced hysteresis. However the 1980s' shock dwarfs previous swings; lending support to the notion that it is the first large (as defined above) shock since the breakdown of Bretton Woods. Figure 3 shows that between 1971 and 1980 the real dollar depreciated relatively steadily, apart from a number of swings which are small compared to the 1980s' swing. Krugman and Baldwin (1987) argue that the 1970s' dollar decline does not reflect a shift in competitiveness (which would be required to induce hysteresis) but rather a bias in foreign productivity growth. Regardless of the cause, Catherine Mann (1986) finds that the U.S. pass-through relationship has been quite stable prior to the 1980s. For these reasons, we test only for hysteresis during the 1980s.

The evidence in this section does not directly test the model. It is simply intended to establish 1) that the historical relationship between the exchange rate and U.S. aggregate, non-oil import prices has shifted in the 1980s, and 2) that the nature of the shift is not inconsistent with the predictions of the model. Mann (1986, 1987a, 1987b), Foster (1986), and Robert Feinberg (1987) find strong evidence of parameter shifts in import price pass-through equations at both the aggregate and industry level. These studies are *ad hoc* in that they contain no formal explanation for the cause, timing, or exact nature of the parameter shifts. Related literature supports the beachhead model by showing that market structure affects foreign firms' pricing behavior (Dornbusch, 1987; Feinberg, 1986; and Mann, 1987a).

A. *Three Testable Implications*

Before turning to the data, we discuss three testable implications of the model. Section II shows that m, and therefore

ε, should jump up upon "announcement" of the overvaluation. The structural break should therefore occur upon announcement. Operationally, we assume that firms realized sometime in the early 1980s that persistent changes in the international mix of fiscal and monetary policies would lead to a prolonged dollar overvaluation. Thus:

Implication 1: A structural break in the pass-through equation should have occurred sometime in the early 1980s.

This implication is robust to the exact details of the imperfect competition model chosen. As long as there is a multiplicity of equilibria and real exchange rate shocks can shift the economy between these equilibria, then implication 1 would hold in the broad class of models where pricing is affected by market structure.

To formulate more specific implications about the structural break, we extend the Section II model to account for lagged exchange rate effects and aggregation. All Section II assumptions are maintained. Additionally we assume foreign firms face delivery lags and stochastic demand curves. Foreign firm j's problem is to

$$(5)\quad \max_{\{j_t^j\}} E\left\{\sum_{t=0}^{\infty} R^t\Big[R^{n_j}P\Big[m_{t+n_j}, y_t^j, v_{t+n_j}^j\Big]\right.$$
$$\left.\times y_t^j - c_t^{*j}e_t y_t^j - G^j\Big]\right\} - F^j,$$

where n_j is its delivery lag and $v_{t+n_j}^j$ is the variety-specific random demand shock. Note that P_{t+n_j} is a function of v and m from $t+n_j$, and y from period t. Adopting the notation that $x_{t|t-1}$ is the expectation of x_t formed at time $t-1$, the typical Euler equation for y_t^j is

$$(6)\quad P_{t+n_j|t}^j = R^{-n_j}\left[\frac{c_t^{*j}e_t}{1-\left(1/\varepsilon^j\left[m_{t+n_j}^h\right]\right)}\right].$$

Here we have assumed a constant elasticity demand curve, where the ε^j is a nonstochastic function of industry h's m. Assuming firms face a variety of delivery lags ranging from 0 to N periods, the log of the aggregate import price index, $P_t = \Pi_{j=1}^{l}[P_{t|t-n_n}^j]^{\theta^j}$, ($l$ is the number of firms, θ^j is firm j's weight in the index) will be related to current and lagged values of an index of firms' marginal costs, $C_{t-i}^* = \Pi_{j\varepsilon\Psi_i}[c_{t-i}^{*j}e_{t-i}]^{\theta^j}$, where Ψ_i is the set of firms with a delivery lag of i periods.

Assuming rational expectations, and using standard macroeconometric arguments, it is easy to derive the time-series properties of the expectational error (u_t) when observed P's are substituted for their expectations. Namely, u_t follows a moving average (MA) process of order $N-1$. Since u is not orthogonal to the regressands, we must instrument. Rationality also implies that any variable lagged N periods or more is a potential instrument.

Empirically we use an import price index, P_t, and a marginal cost proxy, $C_t^{*\prime}$, which we assume are related to P and C^* by $P_t = \nu P_t'$ and $C_t = \delta C_t^{*\prime\rho}$. Defining $\lambda = (\delta/\nu)$, $\Omega^h = \Sigma_{j\varepsilon\Gamma_h} R^{-n_j}$ (Γ_j is the set of firms in industry h), H is the number of industries, and $\varphi_i = (\rho + \Sigma_{j\varepsilon\Psi_i}(\theta^j))$, the pass-through equation is

$$(7)\quad \log(P_t')$$
$$= \log\left(\lambda \prod_{h=1}^{H}\left[\frac{\Omega^h}{1-\left(1/\varepsilon^h\left[m_t^h\right]\right)}\right]^{\theta^h}\right)$$
$$+ \sum_{i=0}^{N} \varphi_i \log(C_{t-i}^{*\prime}) + u_t.$$

The constant term is the parameter we are most interested in. (7) shows the constant term is inversely related to the ε^h's which are directly related to the m_t^h's. According to the beachhead model, the m_t^h's should jump up (for industries with low enough F's), forcing the ε^h's up, and the constant term down. Thus:

Implication 2: The structural break should take the form of a reduction in the constant term of equation (7) beginning sometime in the early 1980s.

In the Chamberlainian framework, additional entry increases the aggregate import price elasticity (which is simply the weighted sum of the variety-specific elasticities). Also each firm has a lower constant term but there are more firms, so the aggregate constant (which is the sum of firm's constants) may increase or decrease. Last, the income elasticity shift is ambiguous. This gives us:

Implication 3: The price elasticity in the aggregate import demand equation should rise (in absolute value) synchronous with the structure break in the pass-through equation. The constant term and income elasticity may rise, fall, or remain unchanged.

B. *Evaluating the Implications*

To test implications 1 and 2, we estimate (7) allowing for a once-off shift in the constant term via the inclusion of an intercept dummy (see the Appendix for data details).[7] Table 1 presents the two-stage least squares (2SLS) estimates of (7) using a WPI-based costs proxy and allowing for a maximum delivery lag of 5 quarters ($N = 5$ was chosen a priori as a reasonable upper bound).[8] Since the exact timing of the break is not known, the equation is estimated for a variety of breakpoints. The results provide strong support for implications 1 and 2. For each breakpoint, the shift in the intercept term is significant and of the expected sign.

The Table 1 results have several problems. First, as is usually the case with pass-through equations, the point estimates of the lagged effects are often negative or insignificant—due most likely to multicolinearity. However, for none of the negative coefficients can we reject the hypothesis that the true parameter is actually positive; nor can we accept the hypothesis that their sum is zero. Next, the Durbin-Watson statistics are low despite the fourth-order MA correction. However, for all but the 1982:1 breakpoint, the values lie within the upper and lower bounds of the 5 percent points. We therefore cannot accept or reject the positive first-order autocorrelation hypothesis.

Although the WPI-based measure is probably a good proxy for marginal costs, it includes only industrialized countries—leaving open the possibility that the shift stems from the exclusion of the newly industrializing countries (NICs). Also, between 10 and 20 percent of U.S. non-oil imports are commodities so the index can be expected to perform poorly during periods of large commodity price swings. Additionally WPIs inevitably contain the price of foreign final goods which do not affect production costs. To partially redress these problems, and to check that our results are robust to the specific-cost proxy, we test implications 1 and 2 on two additional cost proxies: one based on normalized unit labor costs (NULC), and one based on consumer price indices (CPI).

NULC provides a direct measure of a significant component of firms' marginal costs. It avoids the problem of including imported final goods prices, but fails to reflect the cost of imported intermediate goods and commodities (especially fuels). Also it includes only industrialized countries. The third proxy consists of the Federal Reserve Board's foreign costs proxy, which is based on CPIs from the G-10 countries and 8 major NICs. CPIs are a poor proxy for marginal costs (inter alia, they include nontraded and imported final goods prices); however, they are available for a wider range of countries and for longer periods than are more direct measures. Details are in the Appendix.

Table 2 summarizes the results for various breakpoints and N's (multicolinearity hinders precise identification of the lag length). The CPI results, like the WPI results, strongly support implications 1 and 2. The dummies are negative and significant in all cases. The NULC data provide only weak support for the two implications. In only 2 of the 16 cases are the dummies significant at the 90 percent level. However, in both cases they are negative. In all but 4 of the other cases,

---

[7] We assume that the 1980s' shock was a large $V(A,T)$ of the first type since not all competing U.S. firms exited.

[8] For instruments, we used lagged values of the nominal exchange rate, the nominal foreign cost proxy, and nominal domestic costs.

TABLE 1—ESTIMATE RESULTS FOR PASS-THROUGH EQUATION ON WPI DATA, $N=5$

Breakpoint	C	DUMMY	LMC	LMC_{-1}	LMC_{-2}	LMC_{-3}	LMC_{-4}	LMC_{-5}	SE	DW	R^2
1980:3	3.3	−.08	.23	−.38	.14	.11	−.06	.25	.03	1.2	.90
	(10.5)	(−3.3)	(1.1)	(−1.3)	(.8)	(.4)	(−.2)	(1.6)			
1982:1	5.1	−.18	−.20	.04	.22	.01	−.12	−.07	.04	0.9	.80
	(3.8)	(−2.1)	(−.6)	(.2)	(.9)	(.0)	(−.3)	(−.2)			
1982:3	3.9	−.11	−.07	.11	.19	.11	−.13	−.07	.02	1.7	.96
	(10.4)	(−4.2)	(−.4)	(.6)	(2.2)	(.7)	(−.8)	(−.5)			
1983:1	3.3	−.07	−.00	.08	.31	−.11	.04	−0.3	.02	1.3	.95
	(13.4)	(−4.0)	(−.0)	(.3)	(3.0)	(−.5)	(.2)	(−.2)			

Notes:

t-statistics in parentheses.

LMC indicates log of marginal cost proxy.

DUMMY indicates intercept dummy.

C indicates constant term.

SE indicates standard error of regression.

DW indicates Durbin-Watson statistic.

R^2 indicates R-squared statistic.

Sample period 1975:1–1987:1, quarterly data.

TABLE 2—INTERCEPT DUMMIES: VARIOUS DATA SETS AND BREAKPOINTS

		WPI	Unit Labor Cost	CPI
Sample period:		(1987:1–1987:1)	(1987:1–1986:3)	(1967:1–1987:2)
Lag	Breakpoint			
(N = 7)	1980:3	NA	−	−**
	1982.1	NA	−	−**
	1982.3	−**	+	−**
	1983.1	−**	+	−**
(N = 5)	1980:3	−**	NA	−**
	1982:1	−**	+	−**
	1982.3	−**	+	−**
	1983.1	−**	NA	−**
(N = 4)	1980:3	−**	−	−**
	1982.1	−**	−	−**
	1982.3	−**	−*	−**
	1983.1	−**	−	−**
(N = 3)	1980:3	−**	−	−**
	1982.1	−**	−	−**
	1982.3	−**	−**	−**
	1983.1	−**	NA	−**

Notes:

− indicates that intercept dummy is negative, implying that perceived demand elasticity increases in magnitude after breakpoint.

+ indicates that intercept dummy is positive, implying that perceived demand elasticity decreases in magnitude after breakpoint.

NA indicates the test could not be performed due to nonconvergence.

** indicates no change in constant term hypothesis rejected at 95 percent level.

* indicates no change in constant term hypothesis rejected at 90 percent level.

TABLE 3—TESTS FOR TIMING OF STRUCTURAL BREAK IN PASS-THROUGH EQUATION, $N = 5$

	WPI		NULC		CPI	
Breakpoint	Rising Dollar	Falling Dollar	Rising Dollar	Falling Dollar	Rising Dollar	Falling Dollar
1980:3	–*	–**	+	–	+	–**
1982:1	–	–*	–	–	–*	–**
1982:3	–*	–**	–	–	–	–**
1983:1	–**	–**	–	–	–	–**

Notes:

– indicates intercept dummy is negative.
+ indicates intercept dummy is positive.
* indicates dummy is significant at 90 percent level.
** indicates dummy is significant at 99 percent level.

the dummies are negative but insignificant. The insignificance of the NULC estimates may be due in part to the shorter sample period, but also to the nature of the NULC. Since most fuels are priced in dollars, the NULC excludes a cost component which is not directly affected by the exchange rate. Clearly then the NULC underestimates foreign costs during the strong dollar period (1980–1987). This argues that the dummies may be insignificant because the coefficients overestimate the true dummy parameters.

The beachhead model predicts that the parameter shift should have occurred sometime during the rising dollar. Mann (1986) and Foster and Baldwin (1986) suggest the shift occurred at the turning point of the dollar cycle. It is not possible to formally identify the breakpoint from the data; however, it is possible to test these two well-specified hypotheses against each other.[9] To this end we estimate equation (7), with 2SLS, including an intercept dummy for the period from the breakpoint to the peak of the dollar and a second for the post-peak period. If the beachhead model is correct, both dummies should be negative and significant; if the alternative is correct, only the second dummy should be significant. Table 3 presents the results.

For the WPI data (which is arguably the best proxy) the alternative is clearly rejected. For all the various breakpoints both dummies are negative, and in all but one case they are significant. The NULC data tend to reject both hypotheses. In all cases the dummies are insignificant. In three cases both dummies are negative while in one case the first is positive and the second is negative. Again the low degrees of freedom and biases may account for these results.

The CPI data are ambiguous. In one case both dummies are negative and significant. In the other cases the first dummy is insignificant and in one case it is actually positive. The divergence of the CPI and WPI results may be due to the composition of the CPIs. If foreign currency production costs are constant, a rising dollar increases the competitiveness and therefore the sales of foreign firms. However the increased sales tend to put upward pressure on the foreign production costs (for example, wages). Since CPIs place substantial weight on nontraded goods (housing and services) which are not directly subject to the upward pressure, the CPI underestimates foreign production costs during rising dollar periods. Consequently, the point estimates on the first dummy may be insignificant since they underestimate the true parameter.

---

[9] If a test statistic is to have a well-defined distribution under both the null and alternative hypotheses, the model must be correct under both. That is, the hypotheses must be nested. Searching for breakpoints violates this condition. Nevertheless, it may be of interest, that for the WPI data the 1982:3 breakpoint had the highest R-squared.

TABLE 4—STRUCTURAL BREAK TESTS ON IMPORT VOLUME EQUATION, FIRST DIFFERENCES

	Estimation on Subsamples						
Case	Sample Period	*C* (*t*-stat)	GNP (*t*-stat)	Sum of Price Terms (*t*-stat)	*SE*	*DW*	*R*2
1	67:1–87:2	.004 (0.7)	2.07 (4.2)	−0.80 (−2.1)	.04	2.6	.34
2	67:1–80:3	.003 (0.3)	1.91 (2.7)	−0.69 (−1.1)	.05	2.6	.32
	80:4–87:2	.000 (0.0)	2.40 (3.1)	−1.08 (−1.1)	.03	2.7	.35
3	67:1–82:1	.003 (0.3)	1.92 (3.3)	−0.63 (−1.2)	.05	2.6	.32
	82:2–87:2	−.026 (−1.3)	2.93 (2.8)	−2.10 (−1.8)	.03	3.1	.46
4	67:1–82:3	.003 (0.5)	1.89 (3.4)	−0.70 (−1.4)	.05	2.6	.32
	82:4–87:2	−.029 (−1.4)	3.58 (2.8)	−1.63 (−1.3)	.03	2.4	.50
5	67:1–83:1	.003 (0.3)	1.94 (3.4)	−0.67 (−1.4)	.05	2.6	.3
	83:2–87:2	−.021 (−1.7)	2.51 (3.1)	−2.26 (−2.4)	.02	1.6	.73

Estimation with Dummies

Breakpoint	Intercept Dummy	Sum of Price Dummies
1980:3	–	–
1982:1	–	–
1982:3	+	+
1983:1	+	+

Notes:

Price term is sum of coefficients on current and 6 lagged terms.

C indicates constant term.

None of the dummies are significant at the 90 percent level.

Next we turn to implication 3 by testing an aggregate import demand equation for parameter shifts.[10] Specifically, working with first differences, we regress the log of non-oil import volume on the log of real U.S. GNP and the log of current and 6 lags (quarterly data) of relative import prices.[11] The lag coefficients are constrained to follow a third-order polynomial (data details are in the Appendix).

Table 4 presents two types of evidence. The first line shows the point estimates and *t*-ratios for the import demand equation estimated on the entire sample period 1967:1 to 1987:2. The next 8 lines show how they change when the same equation is estimated on pre- and post-break data. For all four breakpoints, the post-break demand is more elastic, providing some support for implication 3. However only in case 5 are the parameter shifts statistically significant. The second type of evidence is the results for import demand equations which include dummies for the constant and price elasticity terms. Slope dummies are allowed for the current and lagged price terms. Only the sum of the lags and dummies are reported.[12] The intercept and slope dummies are insignificant in all four cases and are evenly split between positive and negative point estimates.

Table 4 provides little support for implication 3. This does not necessarily cast doubt on the basic hysteresis hypothesis because implication 3 is not robust to small changes in the market structure assumptions. As is well-known, in a symmetric Cournot equilibrium with homogeneous products, each firm faces the total demand elasticity multiplied by the number of firms. Additional entry makes each firm face a more elastic demand curve but does not shift the aggregate curve. To make our basic theoretical point most clearly, we utilized the familiar Chamberlainian setup. If, as in Baldwin (1986), we had worked with homogeneous products, no volume equation shifts would be expected.

---

[10] A standard assumption used to identify the import demand equation is to assume that supply is infinitely elastic.

[11] Imports and GNP are trended, so the asymptotic distribution of the estimator is not well-defined. First differencing removes the time-series' explosive component.

[12] As usual the point estimates on lagged price effects are often negative or insignificant due most likely to multicolinearity.

Tables 1 and 2 indicate that the elasticities perceived by individual firms increased; but Table 4 indicates that the aggregate elasticity did not. One interpretation of these results is that the most of the new entrants entered markets marked by homogeneous goods.

IV. Conclusions

This paper shows that in a simple industrial organization model, large exchange rate shocks can have persistent real effects, while small shocks cannot. In particular it shows that large exchange rate movements should be correlated with parameter shifts in standard estimated trade equations. Moreover we conjecture that the possibility of hysteresis has theoretical implications for several fundamental issues in international economics. Further theoretical research is required to explore this conjecture.

Empirical tests find evidence for the predicted structural breaks in the U.S. pass-through equation in the 1980s, although not for the import volume equation. In particular, we find evidence that the structural break in the pass-through equation occurred in the rising dollar phase rather than in 1985 as is commonly asserted. Nevertheless, the tests have little power against alternative hypotheses so the breaks may be due to causes totally unrelated to the beachhead model. Further empirical work is needed to directly evaluate the model. Direct tests require time-series data on the number of varieties of imported goods and their close substitutes on an industry level. Unfortunately, such data (or good proxies) appear to exist only for a few industries (for example, automobiles). A direct test of the model on macro data therefore appears to be impossible.

Finally, the beachhead model focuses on supply-side factors, yet empirically demand-side factors may be even more important. In the familiar experience-goods framework, it seems likely that real exchange rate shocks could cause persistent changes in consumer's information sets, and thereby have hysteretic effects. Additional research is needed to explore this conjecture.

APPENDIX

Figure 3 is based on the Federal Reserve Board's foreign cost proxy converted to dollars and divided by the U.S. GNP deflater. The Board's proxy is a trade-weighted index of CPIs from the G-10 and Mexico, Brazil, Singapore, Hong Kong, South Korea, Philippines, and Malaysia. Table 1 uses the NIA non-oil import price and a dollar-denominated, import-weighted marginal costs proxy made from IMF data on manufactured goods WPIs from Canada, Japan, Austria, Belgium, Denmark, Finland, France, Germany, Ireland, Italy, Netherlands, Spain, Sweden, Switzerland, and UK. Both measures are deflated by the U.S.-manufactured goods WPI. For lack of a better measure, we use the NIA deflater (which is a variable weight index based on unit indices). Shifts in import commodity composition may contribute to the parameter shifts. This cannot be the whole explanation since many studies show that even industry-level pass-through equations (where fixed weight indices are used) have shifted.

Table 2 NULC data is a dollar-denominated, import-weighted index of the IMF NULC data from the above-mentioned 15 countries and the NIA deflater, both divided by the U.S. NULC.

Table 4 data are the NIA non-oil import value converted to volume by the corresponding deflater; and the relative price term is the NIA deflater divided by the U.S. GNP deflater. These data choices are standard (Bill Helkie and Peter Hooper, 1986) but subject to well-known measurement error biases. However, there is no reason to suspect that the biases have changed, so the structural break tests are still valid.

REFERENCES

Baldwin, Richard, "Hysteresis in Trade," mimeo., prepared for NBER Summer Institute, International Studies, April 1986.

______, **and Krugman, Paul,** "Persistent Trade Effects of Large Exchange Rate Shocks," NBER Working Paper No. 2017, September 1986.

Bean, Charles, "Sterling Misalignment and British Trade Performance," Centre for Labour Economics, LSE, Discussion Paper No. 288, July 1987.

Chamberlain, Edward, *The Theory of Monopolistic Competition*, Cambridge, MA: Harvard University Press, 1933.

Dixit, Avinash and Stiglitz, Joseph, "Monopolistic Competition and Optimum Product Diversity," *American Economic Review*, June 1977, *67*, 297–308.

______ **and** ______, (1987a) "Hysteresis, Import Penetration, and Exchange Rate Pass-Through," mimeo., November 1987.

______ **and** ______, (1987b) "Entry and Exit Decisions of Firms Under Fluctuating Real Exchange Rates," mimeo., September 1987.

Dornbusch, Rudiger, "Expectations and Exchange Rate Dynamics," *Journal of Political Economy*, December 1976, *84*, 1161–76.

______, "Exchange Rates and Prices," *American Economic Review*, March 1987, *77*, 93–106.

Feinberg, Robert M., "The Effects of Foreign Exchange Movements on U.S. Domestic Prices," ITC mimeo., December 1986.

______, "Recent Foreign Exchange Movements, Prices, and U.S. Trade Flows," ITC mimeo., March 1987.

Foster, Harry, "The Relationship Between U.S. Import Prices and the Exchange Rate: An Empirical Investigation," MIT mimeo., December 1986.

______, **and Baldwin, Richard,** "Marketing Bottlenecks and the Relationship between Exchange Rates and Prices," MIT mimeo., October 1986.

Helkie, Bill and Hooper, Peter, "The U.S. External Deficit: Diagnosis, Prognosis, and Cure," December 1986, in R. C. Bryant and others, eds., *External Deficits and the Dollar: The Pit and the Pendulum*, Washington, D.C., Brookings Institution, 1988.

Krugman, Paul, "Pricing to Market When the Exchange Rate Changes," NBER Working Paper No. 1926, May 1986.

______ **and Baldwin, Richard,** "The Persistence of the U.S. Trade Deficit," *Brookings Papers on Economic Activity*, 1:1987, 1–43.

Mann, Cathy, "Prices, Profits, Margins, and Exchange Rates," *Federal Reserve Bulletin*, June 1986.

______, (1987a) "Prices, Profit Margins and Exchange Rates: After the Fall," DIF Discussion Paper, Federal Reserve Board, July 1987.

______, (1987b) "The Puzzling Behavior of Non-Oil Import Prices," mimeo., November 1987.

Orcutt, Guy, "Measurement of Price Elasticities in International Trade," *Review of Economics and Statistics*, May 1950, *32*, 117–32.

Spence, Michael A., "Product Selection, Fixed Costs, and Monopolistic Competition," *Review of Economic Studies*, June 1976, *43*, 217–35.

The Theory of International Economic Sanctions: A Public Choice Approach

By William H. Kaempfer and Anton D. Lowenberg*

Both economists and political scientists have expressed considerable skepticism about the effectiveness of international economic sanctions as a means to bring about economic dislocation and desired political responses in the target country.[1] It is generally pointed out that sanctions are costly to the sanctioning country as well as the target, that substitution possibilities in world markets diminish the impact of boycotts, and that severe sanctions can induce perverse political responses in the target country whereby its citizens "rally around the flag" and lend increased political support to the ruling regime.

This traditional view of sanctions, however, is premised on an "instrumental" theory, according to which the real intent of sanctions is to bring about policy change in the target nation through imposing the severest possible economic harm. One purpose of this paper is to propose that sanctions might have an altogether different goal —namely, to serve the interests of pressure groups within the sanctioning country. While some of these groups might derive pecuniary benefits from the imposition of sanctions (for example, producers of import-competing goods gain from a ban on imports), others obtain utility directly from taking a moral stance against some other nation's objectionable behavior. Per Lundborg (1987), following Johan Galtung (1959), has labeled this latter motive "expressive," as opposed to instrumental. Especially when it is recalled that trade interruptions which have high costs to the target country are also likely to carry heavy costs for the sanctioning country, it is not surprising that the sanctions actually adopted often appear ineffectual.

However, a second purpose of this paper is to show that this appearance of ineffectuality might be illusory. Sanctions that create minimal economic hardship can still generate political change. If we take the same public choice framework which we use to explain the adoption of sanctions in the sanctioning country and apply it to analyze the political outcome in the target country, it soon becomes clear that sanctions can communicate signals or threats, not necessarily entailing severe economic damage, which in turn produce policy changes. Thus, even sanctions which have little economic impact in the target country can induce desired political responses if they are designed to selectively affect the appropriate interest groups. Much of the traditional wisdom on sanctions implicitly assumes that their intended purpose is instrumental—that their real goal is to impose maximum economic harm in the target country. Thus, for example, Hufbauer and Schott (1985), maintain that successful sanctions episodes are generally those in which heavy costs are imposed on the sanctioned nations. The analysis developed in the present paper suggests this might not be so. Nevertheless, our conclusions *are* consistent with another of Hufbauer and Schott's generalizations—namely, that sanctions which are very costly to the sanctioning country are unlikely to succeed, precisely because, according to our analysis, such sanctions are often designed deliberately to be ineffectual.

We use an equilibrium model of interest group competition within the sanctioning country and the target country, in which public policy outcomes are explained as a function of the relative effectiveness of interest groups in generating political pressure. One purpose of the model is to predict the content and level of sanctions imposed by the sanctioning country. For instance, this approach might be used to explain why sanctioners typically restrict imports from the target country rather than exports to that country, or to explain the types of imports singled out for embargo. We use the same model to explain the political response of the target country, and to account for the political effectiveness of sanctions with weak economic effects.

---

*Department of Economics, University of Colorado, Boulder, CO 80309-0256, and Department of Economics, California State University-Northridge, CA 91330. We thank James Alm, Thomas Hazlett, Ron Heiner, James Lehman, Gordon Tullock, Thomas Willett, and two anonymous referees for helpful comments.

[1] See, for example, Klaus Knorr, 1975, Donald Losman, 1979, Margaret Doxey, 1980, Thomas Bayard, Joseph Pelzman, and Jorge Perez-Lopez, 1983; and Gary Clyde Hufbauer and Jeffrey Schott, 1985.

I. A Model of the Political Process in the Sanctioning Country

The level of economic sanctions, as an international public policy tool, is determined by pressures brought to bear in the political system by interest groups of differing motives. These pressures are the outcome of private utility maximization on the part of individual members of the interest groups concerned. While the groups are defined by commonality of interests, the political participation of any group member is tempered by a desire to free ride.

Consider some individual i, a member of the total population I of some state. This individual maximizes utility according to

$$(1)\quad \operatorname{Max} U^i = U^i(Y^i),\quad U_1^i > 0,\quad U_{11}^i < 0$$

$$\text{subject to}\quad Y^i = Y^i(S),\quad Y^i(0) = E^i,$$

$$Y_1^i \gtrless 0,\qquad Y_{11}^i = 0,$$

where Y is income, E is an initial endowment, and S is a nonnegative, continuous variable measuring the level of sanctions applied against some external state. These sanctions are essentially specific instruments of protection that regulate goods or factor flows with the target nation. Given this type of regulatory intervention in the market, sanctions increase or decrease individual income at a fixed rate, as a simplifying assumption. Since these sanctions are interventions that generate distortions, $\Sigma_i Y_1^i < 0$. That is, they are inefficient, although they will increase the income of particular interest groups at the expense of others.

The change in utility produced by a change in the level of sanctions is

$$(2)\qquad \delta U^i/\delta S = U_1^i \cdot Y_1^i \gtrless 0.$$

Let $I = \{J, K, L\}$ such that for all $j, k, l \in J, K, L$, respectively, $Y_1^j > 0$, $Y_1^k < 0$, $Y_1^1 = 0$. This allows separate demand functions for sanctions to be specified for those who benefit from sanctions and for those who suffer from sanctions. For group J, all members' willingness to pay for an additional unit of sanctions can be summed to yield

$$(3)\qquad P_S = D^J(S) = \sum_j \delta U^j/\delta S = \sum_j U_1^j \cdot Y_1^j,\qquad D_1^J < 0,$$

where P_S is the unit price of sanctions. This is the dollar amount that individuals in group J would be willing to spend to achieve a certain level of utility-enhancing sanctions through the political market. This demand curve is downward sloping due to the specification of equation (1).

Similarly, members of group K are willing to pay to avoid sanctions that could be imposed through the political market because $Y_1^k < 0$. This implies a demand for reduced sanctions, which, when expressed as a function of an increasing level of sanctions, becomes

$$(4)\qquad P_S = D^K(S) = -\sum_k \delta U^k/\delta S = -\sum_k U_1^k \cdot Y_1^k,\qquad D_1^k > 0.^2$$

---

[2] This expression is defined as the negative of the sum of the marginal utilities of the change in Y from a change in S. Essentially, group K loses utility as sanctions grow. We interpret this, at the margin, as the amount that group K members would pay to keep sanctions from increasing.

The demand for sanctions must take into account the public good nature of this type of market intervention. The demand functions (3) and (4) are derived by summing the maximum willingness to pay for more sanctions or fewer sanctions over the members of groups J and K. In an attempt to free ride on other group members, however, each member has an incentive to reveal a lower willingness to pay. The extent to which this ability to free ride diminishes the demand of a group depends on the political effectiveness of that group. This suggests a respecification of demand,

$$(3') \quad P_S = J(S, E^J), \quad J_1 < 0, \quad J_2 < 0,$$

$$(4') \quad P_S = K(S, E^K), \quad K_1 > 0, \quad K_2 < 0,$$

where E^J and E^K are shift parameters which reflect the degrees of free riding in groups J and K, which in turn determine the abilities of the two groups to produce political influence (Gary Becker, 1985). Free riding is a function of the size of the group and other factors which influence organization and enforcement costs (James Buchanan and Gordon Tullock, 1962). The more severe the free-riding problem, the greater the magnitude of E^J and E^K.

Equating the demand for more sanctions, (3) or (3′), with the demand for fewer sanctions, (4) or (4′), will clear the political market. The government acts to clear the market by raising the level of sanctions above its zero minimum until pressure for higher sanctions is offset by pressure against higher sanctions. Figure 1 depicts such a political market for sanctions, given the following additional assumptions. First, assume initially that there are no output deadweight losses from sanctions. Thus the sanctions are merely lump-sum redistributional tools, that is, $\Sigma_i Y_1^i = 0$. Second, assume the marginal utilities of all individuals in the neighborhood of $S = 0$ are identical. Thus, for an infinitesimally small level of sanctions, since the income gain of group J is assumed exactly equal to the loss of group K, it follows that the willingness to pay of group J for more sanctions is equal to the willingness to pay of group K for less sanctions. The political market therefore clears at $D^K = D^J \Rightarrow S = 0$.

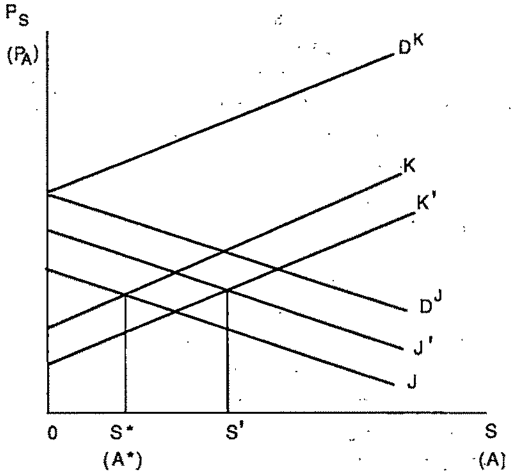


FIGURE 1

An increase in sanctions will be positively valued by group J along D^J, but at a decreasing rate because the marginal utility of Y is assumed to be decreasing. Similarly, group K is willing to pay increasing amounts to prevent sanctions from rising, again due to the rising marginal disutility of falling Y.

More typically, the market-clearing level of sanctions would be zero due to the nonnegative lower bound on S. For instance, the presence of deadweight costs, implying that $\Sigma_i Y_1^i < 0$, would lead group K to evaluate its loss from lower income due to any positive level of sanctions at a greater amount than the value of the income gain to group J. Alternatively, if the government has an administrative cost of applying sanctions, the pressure for sanctions not only has to cover the pressure against sanctions, but also pay for the administrative costs of the sanctions. In both cases, this amounts to a downward shift of D^J with respect to D^K, resulting in an equilibrium at a negative level of sanctions—precluded by the assumption of nonnegativity of S.

Without differences in the political effectiveness of various interest groups, sanction interventions are unlikely to be applied by a government responding only to income-maximizing political pressure. As Gary Becker (1985, p. 344) points out, "... no policy that lowered social output would survive if all groups were equally large and skillful at producing political influence, for the opposition would always exert more influence than proponents." However, the political ineffectiveness of groups J and K will cause their respective demand curves, $J(S, E^J)$ and $K(S, E^K)$, to shift below D^J and D^K. If J is a small group, each member of which gains a significant share of the increase in income obtained by an increase in sanctions (that is, manufacturers of substitutes for the exports of the target country), then J may be much more politically effective than group K (comprising the consumers who lose from import restrictions). Consequently, if E^J is much smaller than E^K, the downward shift of $J(\cdot)$ below D^J might be small enough relative to the downward shift of $K(\cdot)$ below D^K to make sanctions politically efficient as shown in Figure 1 at S^*.

In general, however, political pressure for economic sanctions will rise not only due to the income effects of sanctions in terms of increased consumption possibilities for members of certain interest groups, but also due to utility enhancing attributes of the sanctions themselves.[3] That is, consider sanctions a public good (or bad) that directly contributes to individual utility (or disutility) by allowing individuals the satisfaction (or dissatisfaction) of experiencing their nation engaged in a foreign policy toward a certain goal. This implies a revised utility-maximization process based on a respecified utility function

$$(1') \qquad U^i = U^i(Y^i, S).$$

Differentiating this expression with respect to S implies a direct and indirect effect on utility from imposing sanctions

$$(5) \qquad \delta U^i/\delta S = \delta U^i/\delta S|_{Y_1^i=0} + U_1^i \cdot Y_1^i.$$

Assume initially the direct effect,

$$\delta U^i/\delta S|_{Y_1^i=0},$$

is positive. This implies that the pressure brought to bear by the pro-sanctions group J will increase, thereby shifting the demand for sanctions to J' in Figure 1. At the same time, although group K loses income from sanctions, its members obtain direct positive utility from the imposition of sanctions. This implies a downward shift to K' in Figure 1. Thus when all members of I derive utility directly from sanctions, the level of sanctions will rise to S'.

However, it could be the case that while some members of society gain direct utility from sanctions, others suffer a direct diminution in utility from the imposition of sanctions quite apart from any income effects. Consider a third group L for whom $Y_1^1 = 0$ but $\delta U^1/\delta S \gtrless 0$. This implies a process of political strategies and alliances between groups J and K and portions of group L that augments the pressures of those whose income is directly affected by sanctions. In general both curves J and K would shift upward in Figure 1 and the political market-clearing level of sanctions would depend on the relative intensities of preference and political inefficiencies in the pro- and anti-sanctions portions of L.

Thus the above analysis shows that the actual level of sanctions imposed by the sanctioning country is a function of the relative political influence of interest groups within that country, and that the equilibrium level of sanctions which is produced from the political market-clearing process is politically efficient. This implies not only that the level or composition of sanctions which ultimately prevails will be income reducing in total for the sanctioning country, but also that there is no reason why it should necessarily coincide with that level of sanctions required to induce maximum economic harm in the target country.

---

[3] In addition to the "instrumental" motive of imposing economic damage on the target country, embargoes can have an "expressive" value to sanctioners who feel obliged to act for moral or political reasons (Per Lundborg, 1987). In the absence of such effects on utility, sanctions are analytically identical to country-specific protection.

If the purpose of sanctions is to placate domestic pressure groups interested in expressing their disapproval of the practices of a foreign government, this might be achieved with minimal deadweight costs to the sanctioning economy by using sanctions with minimal market effects (Thomas Willett and Mehrdad Jalalighajar, 1983/84). This can be done by applying sanctions against countries with which the volume of trade is small, or by targeting markets in which the degree of substitutability between alternative suppliers of goods is high. Also, it is to be expected that those interest groups in the sanctioning country which are more efficient producers of political influence will succeed in harnessing the sanctions strategy to their own benefit. It is likely that industries seeking protection from foreign imports will ally themselves with other interest groups seeking to sanction the exporting nation. Thus U.S. sugar producers would presumably support sanctions against Cuba, and it is surely no coincidence that the package of U.S. sanctions against South Africa (implemented in 1986) includes restrictions on imports of textiles and steel, which industries are also important sources of protectionist pressure. Our approach also suggests that, since producers are a more cohesive and politically effective interest group than consumers, sanctions are more likely to restrict imports from the target country than exports to the target.

II. The Optimal Sanction for Target Country Policy Change

The polity of the target country also comprises two primary interest groups. One of these exerts pressure through the political market in favor of a policy package, which we will designate A, and which is deemed objectionable by pro-sanctions interest groups in the sanctioning country. The other interest group lobbies against A. The model used in this section is identical to that developed above, except that the endogenous policy outcome which emerges from the political market-clearing process is the level of A instead of the level of sanctions. This policy package, A, supplied by the state, consists in a series of regulations which effect redistributions between domestic interest groups. In aggregate, these regulations are income decreasing. We will again assume that the population I comprises two major interest groups—J, whose income is increased with higher levels of A, and K, whose income is reduced by A.

The political market-clearing mechanism in the target country is described by equations (1)–(4) and Figure 1, with the objectionable policy A replacing S. Again, the curves J and K in Figure 1 are drawn to reflect the assumption that group J (the beneficiary of A) is more effective at producing political influence than group K. This will result in a positive level of the optimal policy outcome, A^*. The relative political ineffectiveness of group K might occur if that group is constitutionally excluded from the nominal policymaking process. The only options open to members of such a group are costly ones like insurrection or civil disobedience, which reduces their revealed willingness to pay for lower levels of A.

Now, assume that economic sanctions imposed on the target country by foreigners are income reducing for members of both interest groups.[4] The individual budget constraint must be rewritten as

$$(6)\quad Y^i = Y^i(A,S), \qquad Y_1^i \gtrless 0, \qquad Y_2^i < 0.$$

Thus sanctions have an indirect negative effect on utility which works through income

$$(7)\qquad \delta U^i/\delta S = U_1^i \cdot Y_2^i < 0.$$

---

[4] In general this need not be true. For instance, see Richard Porter (1979) for a more detailed examination of the differential effects of sanctions on income groups within the target country. Sanctions restricting imports of the target country might increase the market power of producers of import substitutes. So too, disinvestment sanctions can bring about a fire-sale selloff of foreign-owned domestic assets leading to a wealth transfer to those target country residents who acquire them at depressed prices (William Kaempfer and Anton Lowenberg, 1986).

Also, assume that sanctions, in addition to their effect on income, are a determinant of the political ineffectiveness of groups J and K:

$$(8) \qquad E^{J}=E^{J}(S), \qquad E_{1}^{J} \gtrless 0,$$
$$E^{K}=E^{K}(S), \quad E^{K}I_{1}<0.$$

The functional relationships between the political ineffectiveness parameters and sanctions capture the varying perceptions on the part of individual group members of the moral message or signal conveyed by sanctions. Members of group J, who are supporters of the regime which supplies the public good A, might regard sanctions as a foreign interference with national sovereignty, in which case they will "rally around the flag" and the group as a whole will expend more resources on the production of political influence in favor of the policy A. In effect, the incentive to free ride within group J will diminish and $E_{1}^{J}<0$. Alternatively, if individual members of group J are discouraged in their support for A by the income-reducing impact of sanctions, their incentive to free ride in the production of pro-A pressure will increase and $E_{1}^{J}>0$. Members of the opposing group K, however, might regard the imposition of sanctions as a signal of foreign support for their struggle against the A-producing government, in which case group K will become more efficient in exerting political pressure and free riding will diminish, so that $E_{1}^{K}<0$.

Figure 2 shows the possible effects of sanctions on the optimal level of A. The income effects of sanctions are given by the demand functions D_{1}^{J} and D_{1}^{K}. Both interest groups experience a decrease in income, which reduces the demand price revealed by each, either for the public good A (in the case of J) or for reductions in A (in the case of K). Both demand curves have therefore shifted down from their original positions (D^{J} and D^{K}), and the effect on A^{*} is indeterminate since it depends on the relative magnitudes of the shifts. In Figure 2, A^{*} is unchanged. Clearly, however, it is conceivable that the income effect of sanctions might well entail an increase in A^{*}, if group K experiences a greater reduction in income than group J. This implies that, in order to be successful in producing a diminution of the objectionable policy of the target country, sanctions should be designed so as to hurt the prime beneficiaries of that policy to a greater extent than those groups opposing it.[5]

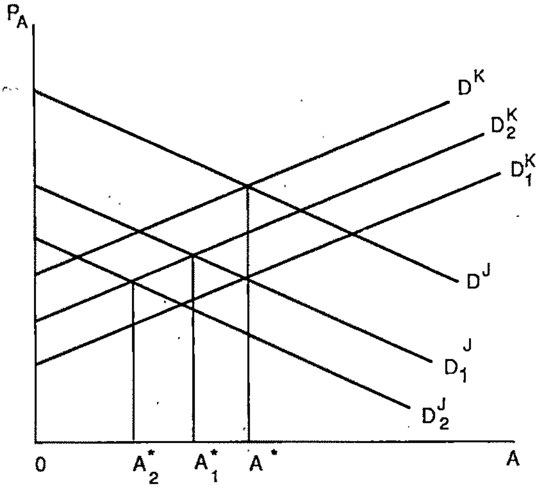


FIGURE 2

If sanctions signal the support of foreigners for the struggle against A, the parameter E^{K} will decrease and the demand schedule of group K will shift up to D_{2}^{K}, leading to a fall in A^{*} to $A^{*}{}_{1}$. As Gordon Tullock (1971) has pointed out, anything which increases the probability of successful political resistance or lowers the expected costs to individuals of political participation will lead to an increase in resistance activities. Sanctions might serve to transmit a signal to the opposition group K that they have the support of foreign governments and interest groups, and therefore reduce their incentive to free ride in the collective production of political pressure directed against policy A. Furthermore, if sanctions are applied incrementally, holding out the threat of further sanctions in the future if the level of A is not reduced (or if it is promised that existing sanctions will be lifted in the future if policy A is retrenched), then a "discouraged rent seeker" effect on members of group J might more than offset the "rally around the flag" effect.[6] In other words, individuals in group J will perceive further pursuit of the policy A as increasingly futile and wealth decreasing, so that they will have an increased incentive to free ride in the production of pro-A political influence and E^J will increase. This group's demand curve will shift down to D_2^J, leading to a further reduction in the optimal level of A to A^*_2.

---

[5] However, if A is an inferior good to group J, then the negative income effect will induce an *upward* shift of D^{J}, leading to an increase in A^{*}. For example, sanctions that reduce the incomes of white workers in South Africa might lead to an increase in the demand for regulations protecting them from black labor market competition, which translates into an increased "demand" for apartheid (William Kaempfer, James Lehman, and Anton Lowenberg, 1987).

Thus, signal or threat effects of sanctions operate to bring about changes in A^* which are quite independent of the impact of sanctions on the income of the target country. Sanctions which generate the greatest economic damage need not necessarily produce a decrease in A^*, and might even precipitate an increase in A^*. A large negative income effect will cause a large downward shift of curves D^J and D^K in Figure 2, but unless group J is hurt proportionally more than group K so that D^J shifts down by more than D^K shifts down, there will be no reduction in A^*. Thus only sanctions which have differential impacts on the supporters and opponents of A are able to produce policy change in the target country. One way to ensure differential impacts is to design the sanctions strategy with a view to its signal and threat effects rather than just its straightforward income effects. A signal of support for the goals of the political opposition, or threats of further pain in the future to the supporters of the regime, will alter the political inefficiency parameters applying to these two interest groups and thereby produce endogenous policy changes in the desired direction.

III. Implications and Conclusions

The configuration of interest groups and their differential abilities to generate political pressure within the sanctioning country ensure that any trade or investment restrictions enacted reduce income in that country due to economic distortions, but are unlikely to be designed specifically to impose the maximum amount of economic dislocation on the target country. Furthermore, the structure of interest group politics within a given country helps determine whether that country will have a high propensity to resort to sanctions instead of other foreign policy tools, and also which target nations it will single out to be at the receiving end of its wrath. The type of foreign economic policy that is implemented depends crucially on its real (as opposed to stated) goal. Thus the types of sanctions observed are likely to be those which serve the interests of influential pressure groups within the polity of the sanctioning country, and are unlikely to be designed in order to impose maximum harm on the target country. Nevertheless, this does not mean that sanctions do not work.

The sanctions which are most likely to precipitate the desired political change in the target country are those which concentrate income losses on groups benefiting from the target government's policy, those which signal political support to opposition interest groups, or those which threaten increased pain in the future and therefore create an incentive for individual supporters of the target government to free ride on the political activities of their group as a whole. Thus, from the point of view of the sanctioning country, those interest groups desirous of achieving political change in the target nation might rationally ally themselves with other interest groups desirous of imposing sanctions for narrow income-enhancing reasons (for example, protectionists), even if the sanctions ultimately enacted are not particularly damaging in economic terms. However, we cannot exclude the possibility that a sanction with little economic impact on South Africa, for example, might be conducive to ending apartheid.

---

[6] An example is provided by Ronald Findlay and Mats Lundahl (1987) who argue that white South African workers might respond to divestment pressure from shareholder groups abroad by conceding to a diminution of apartheid in order to prevent erosion of their standard of living.

The advantage of the public choice approach is its treatment of public policy as an endogenous outcome of competitive interest group pressures within the sanctioning and the target countries. It is thereby revealed that economic harm per se is not the crucial consideration in determining the nature of the sanctions package implemented by the sanctioning country. Economic damage is also not the only, or even the most important, dimension of the effect of sanctions on the target country.

REFERENCES

Bayard, Thomas O., Pelzman, Joseph and Perez-Lopez, Jorge, "Stakes and Risks in Economic Sanctions," *The World Economy*, March 1983, *6*, 73–87.

Becker, Gary S., "Public Policies, Pressure Groups and Dead Weight Costs," *Journal of Public Economics*, December 1985, *28*, 329–47.

Buchanan, James M. and Tullock, Gordon, *The Calculus of Consent*, Ann Arbor: University of Michigan Press, 1962.

Doxey, Margaret P., *Economic Sanctions and International Enforcement*, 2nd ed., New York: Oxford University Press, 1980.

Findlay, Ronald and Lundahl, Mats, "Racial Discrimination, Dualistic Labor Markets and Foreign Investment," *Journal of Development Economics*, forthcoming 1987.

Galtung, Johan, "Pacifism from a Sociological Point of View," *Journal of Conflict Resolution*, March 1959, *3*, 67–84.

Hufbauer, Gary Clyde and Schott, Jeffrey J., *Economic Sanctions Reconsidered: History and Current Policy*, Washington: Institute for International Economics, 1985.

Kaempfer, William H., Lehman, James A. and Lowenberg, Anton D., "Divestment, Investment Sanctions, and Disinvestment: An Evaluation of Anti-Apartheid Policy Instruments," *International Organization*, Summer 1987, *41*, 457–73.

______, ______ **and Lowenberg, Anton D.,** "A Model of the Political Economy of International Investment Sanctions: The Case of South Africa," *Kyklos*, 1986, Fasc. 3, *39*, 377–96.

Knorr, Klaus, *The Power of Nations: The Political Economy of International Relations*, New York: Basic Books, 1975.

Losman, Donald L., *International Economic Sanctions: The Cases of Cuba, Israel, and Rhodesia*, Albuquerque: University of New Mexico Press, 1979.

Lundborg, Per, *The Economics of Export Embargoes: The Case of the U.S.-Soviet Grain Suspension*, London: Croom Helm, 1987.

Porter, Richard C., "International Trade and Investment Sanctions: Potential Impact on the South African Economy," *Journal of Conflict Resolution*, December 1979, *23*, 579–612.

Tullock, Gordon, "The Paradox of Revolution," *Public Choice*, Fall 1971, *11*, 89–99.

Willett, Thomas D. and Jalalighajar, Mehrdad, "U.S. Trade Policy and National Security," *Cato Journal*, Winter 1983/84, *3*, 717–27.

Increasing Returns, Public Inputs, and International Trade

By SUEZO ISHIZAWA*

Governments often finance the development of new technologies. Such new technologies are public inputs whose simultaneous use by several industries imparts a kind of increasing returns to scale to the economy. By means of an example incorporating such a government-financed public input, this note demonstrates that (*i*) differences in the absolute amounts of factor endowments *alone* can cause trade, and (*ii*) such a trade can exhibit Leontief's paradox.

Consider an economy which uses its endowments of labor L and capital K to produce two tradable private goods X_1 and X_2 and a public input X_0. The public input is not traded internationally but is purchased by the government and made available to the private-good industries. Their production functions are given by[1]

$$(1) \qquad X_0 = K_0/2,$$

$$(2) \qquad X_1 = (X_0+1)\min(L_1, K_1/5),$$

$$X_2 = (X_0+1)\min(L_2, K_2).$$

The public input thus raises the output of each private good by equal proportions. We call this the *Symmetry Assumption*. Moreover, a proportionate expansion in the economy's endowments of L and K can result in a more than proportionate increase in the output of each private good; increasing returns to scale prevail at an economywide level. Let L^p and K^p denote labor and capital used for private goods production. For endowments to be fully used, their ratio k^p $(\equiv K^p/L^p)$ and the traditional capital-labor ratio k $(\equiv K/L)$ must satisfy the following condition[2]

$$(3) \qquad 2X_0/L + k^p = k.$$

Commodity and factor markets are assumed to be perfectly competitive. Let w and r denote the price of labor and capital, respectively, and $c^0(w,r)$ the unit or marginal cost of the public input. Its supply is government financed to the point where its marginal benefits are equal to its marginal cost, that is,

$$(4) \quad p_1 \partial X_1/\partial X_0 + p_2 \partial X_2/\partial X_0 = c^0(w,r),$$

where p_1 and p_2 are market prices for X_1 and X_2. Given the supply of the public input, the unit cost of each private good $c^i(X_0, w, r)$ is obtained from (2). The zero-profit condition under perfect competition requires that the price of each private good be equal to its unit cost

$$(5) \qquad p_i = c^i(X_0, w, r), \qquad i = 1,2.$$

Equations (3), (4), and (5) can be solved for (w, r, k^p, X_0) in terms of (p_1, p_2, L, k). Let $\pi(\equiv p_1/p_2)$ and $x(\equiv X_1/X_2)$ denote the relative price and relative output of commodity 1, respectively. Then[3]

$$\begin{aligned}(6) \qquad k^p &= \hat{k}^p(\pi, L, k)\\ &= [(k/2 + 1/L)\\ &\qquad -(5-\pi)/2(\pi-1)],\end{aligned}$$

$$(7) \quad x = x^s(k^p) = (k^p - 1)/(5 - k^p).$$

---

*Department of Economics, University of Missouri-Columbia, Columbia, MO 65211. This paper is based on my unpublished doctoral dissertation submitted to The Johns Hopkins University, 1986. I express my gratitude to my advisors Tatsuo Hatta and M. Ali Khan. Particularly, I am deeply indebted to Tatsuo Hatta, without whose help and encouragement the thesis would not have been completed. I am also grateful to R. Manning, my colleagues C. Menezes and P. Mueser, and an anonymous referee for helpful suggestions.

[1] Although these specifications are highly restrictive, all the results in this note hold for neoclassical production functions. See Suezo Ishizawa, 1986.

[2] From (1) the resource constraints are $L^p = L$, $2X_0 + K^p = K$. These yield (3).

[3] Let $\omega = w/r$. From (5) $\omega = (5-\pi)/(\pi - 1)$; from (4) $\omega L^p + K^p = 2(X_0+1)$. These and the resource constraints in fn. 2 yield (6). Let $Y_i \equiv X_i/(X_0+1)$, $i = 1,2$.

The "supply function" of the relative output of commodity 1 is obtained by evaluating (7) at $\hat{k}^p$ in (6); it is upward sloping against its own price even though increasing returns are imbedded in the economy.[4] Furthermore, the supply function shifts inward for a proportionate expansion in the economy's endowments of L and K. Thus for a homothetic social welfare function, a proportionate expansion of the economy will result in a higher autarkic price of the capital-intensive commodity 1. These results produce the following proposition.

PROPOSITION: *Two countries, each producing all three goods under the same technology, the same homothetic preferences, and the same capital-labor endowment ratio, will trade if they differ in size. The large country will export the labor-intensive commodity, and continue to do so even if there is a slight increase in its capital-labor endowment ratio;*[5] *hence such a trade exhibits Leontief's paradox.*

The basis for this proposition is that in our model the traditional capital-labor ratio k does not regulate the economy's supply function directly, but bears upon it only indirectly via its impact upon $\hat{k}^p(.)$. R. Manning and John McMillan (1979) have developed an elasticity concept which plays the key role in determining how a change in k affects $\hat{k}^p(.)$. Specifically, Ishizawa (1986) has shown that under the Symmetry Assumption, a proportionate expansion in endowments *lowers* (raises) $\hat{k}^p(.)$ if and only if the derivative of $\varepsilon(X_0) = \ln X_i/\ln X_0$, $i=1,2$, with respect to X_0 is *positive* (negative). In the present context, $\varepsilon(X_0)$ equals $X_0/(X_0+1)$ and has a positive derivative. Hence, as the economy expands proportionately, $\hat{k}^p(.)$ falls, and this falling $\hat{k}^p(.)$ causes trade and Leontief's paradox. On the other hand, if $\varepsilon(X_0)$ has zero derivative, as in Makoto Tawada and Kenzo Abe (1984), $\hat{k}^p(.)$ will be independent of country size, and as shown by them, the Heckscher-Ohlin theory holds for such a case.

Our analysis strengthens the results obtained by Manning and McMillan (1979); they introduced a public input into the Ricardian model of one primary factor and showed that trade would not occur under the Symmetry Assumption. We have demonstrated[6] that when there are two primary factors of production, (i) trade can occur even under the Symmetry Assumption if two countries differ in size, but (ii) such trade can exhibit Leontief's paradox.

From this and (2), we have $Y_1 + Y_2 = L^p$, $5Y_1 + Y_2 = K^p$. These and $Y_1/Y_2 = x$ then yield (7). Here we assume full use of L^p and K^p, which holds when their ratio k^p lies between the capital-labor ratios of two private-good industries. This holds for a certain set of commodity prices and factor endowments as seen from (6).

[4] This upward-sloping curve is not a mere incidence of our example. It holds under the Symmetry Assumption. See Ishizawa, 1986.

[5] Let economy I's endowments be $(L_I, K_I) = (1,10)$, and economy II's $(L_{II}, K_{II}) = 5(1, 10+81/185)$, each with the same homothetic social welfare function $X_1^{5.4}X_2$. Then the large, capital-abundant economy II faces a relatively higher autarkic price of the capital-intensive commodity 1, $\pi*_I = 1.8$, $\pi*_{II} = 2$; thus economy II imports that commodity in trade. Such trade exhibits Leontief's paradox.

[6] Our results conform to those of Manning and McMillan when the capital-labor ratio is the same in all industries (i.e., zero). This was pointed out by a referee.

REFERENCES

Ishizawa, Suezo, "Public Inputs, Increasing Returns, and International Trade," unpublished doctoral dissertation, The Johns Hopkins University, 1986.

Manning, R. and McMillan, J., "Public Intermediate Goods, Production Possibilities, and International Trade," *Canadian Journal of Economics*, May 1979, *12*, 243–57.

Tawada, Makoto and Abe, Kenzo, "Production Possibilities and International Trade with a Public Intermediate Good," *Canadian Journal of Economics*, May 1984, *17*, 232–48.

Credibility and Policy Convergence in a Two-Party System with Rational Voters

By ALBERTO ALESINA*

The traditional approach to modeling political parties' behavior, based upon the contribution of Anthony Downs (1957), assumes that the parties' unique objective is to win elections: thus, they maximize their popularity. The crucial implication of this assumption for a two-party system is that if the two parties have the same information about voters' preferences, full convergence of policies results from electoral competition. This is the crucial implication of the "median voter theorem."[1]

More generally, it may be argued that different parties are motivated differently because they represent different constituencies. Parties may not care only about winning elections per se, but also about the quality of the policies resulting from an election. In this case the candidates of the two parties view winning an election not only as a goal per se, but also as a means of implementing a better policy for their respective constituencies.

This paper shows that electoral competitions imply dynamic inconsistency if the voters are modeled as rational and forward-looking agents and parties do not care exclusively about being elected, but also about which policy to implement, once elected. The dynamic inconsistency arises as follows: the parties have an incentive to announce convergent platforms to increase their chances of election. However, if the elected party is not committed to its electoral platform, it has an incentive to follow its most preferred policy rather than the policy announced in its platform. If voters are rational, they account for this incentive. Thus, in general, in a one-shot electoral game the only time-consistent equilibrium is one in which no convergence is possible, the two parties follow their most preferred policies, and the voters rationally expect this outcome. Full convergence of parties' platforms results only as a limiting case when the parties are completely indifferent with respect to the quality of the policies resulting from the election. Thus, these results differ from the existing literature on "ideologically motivated" politicians (for instance, Donald Wittman, 1977, 1983; Randall Calvert, 1985), which implicitly assumes the possibility of binding commitments to electoral platforms.

Complete or partial policy convergence can be the outcome of political competition if the interaction between the parties and the voters is modeled as an infinitely repeated game. In fact, if the candidates have concave objective functions, the welfare-maximizing policy rule implies a complete convergence of parties' policies. However, this cooperative, and agreed-upon policy, may or may not be sustainable as a subgame-perfect equilibrium depending on parameter values; in particular it depends on the discount rates of the two parties, the degree of polarization of their preferences, and the relative popularity of the two parties. For the case in which the first-best policy is not sustainable, the set of credible policies are also characterized as a function of the same variables.

---

*Graduate School of Industrial Administration, Carnegie Mellon University, Pittsburgh, PA 15213, and National Bureau of Economic Research, Cambridge MA, 02138. This paper is based upon a chapter of my unpublished doctoral dissertation at Harvard University. I am greatly indebted to Jeffrey Sachs for directing my attention toward these issues and for many conversations. I also wish to thank Andrew Abel, Dilip Abreu, Olivier Blanchard, Ramon Caminal, Andrew Caplin, Alex Cukierman, Morris Fiorina, Benjamin Friedman, Herschel Grossman, Howard Rosenthal, and the referees for very useful comments. The responsibility of any mistakes is, of course, only mine.

[1] The result of policy convergence in a two-party system is more general than the "median voter theorem." For discussions of convergence results not at the median, see John Ledyard, 1984; Peter Coughlin, 1984; Coughlin and Shmuel Nitzan, 1981; Melvin Hinich, 1977. For earlier work on spatial competition see Richard McKelvey, 1975; Hinich, Ledyard, and Peter Ordeshook, 1972, 1973, and the references quoted therein. The present paper focuses on the result of convergence rather than on the "median voter theorem" per se.

Thus, this paper shows how the repeated interaction of political parties can reduce the magnitude of policy fluctuations. This smoothing effect can be particularly beneficial if excess volatility, for example in economic policy, is harmful for society as a whole. An example related to monetary policy is developed in Alberto Alesina, 1987, and empirically tested in Alesina and Jeffrey Sachs, 1988, and Alesina, 1988.

I. The Model

Let us consider a political system with two parties, denoted party 1 and party 2. No distinction is made between "candidates" and "parties": a candidate is completely identified by the objective function of his party.[2] The parties are not concerned solely about winning elections but have also preferences defined on policy outcomes, since they represent constituencies with different preferences.

Party 1's objectives are represented by the following unidimensional function

$$(1)\qquad U(l)=\sum_{t=0}^{\infty}q^t u(l)=-\sum_{t=0}^{\infty}\tfrac{1}{2}q^t(l_t-c)^2;\qquad c>0\quad 0<q<1.$$

Party 2's objective function is the following:

$$(2)\qquad V(l)=\sum_{t=0}^{\infty}q^t v(l)=-\sum_{t=0}^{\infty}\tfrac{1}{2}q^t l_t^2.$$

The two parties have different bliss points on the policy variable l, which is directly controlled by the party in office. Party 1's bliss point is c and, without loss of generality, party 2's bliss point is normalized at zero. The quadratic specification for (1) and (2) is adopted purely for simplicity; all the results of the paper generalize to any single-peaked, concave objective function. The discount factor (q) is identical for the two parties. Each party attributes a positive utility, denoted by k, to being in office per se, in addition to their objective function defined on l.

The voters have preferences defined on l and they are rational, forward looking, and informed about the objective functions of the two parties (i.e., they know the objective functions (1), (2), and k). Thus, the voting decisions are based upon the rational expectations of the policies that the two parties would follow if elected. Let us indicate these policies with x for party 1 and with y for party 2: if party 1 is elected at time t, it sets $l_t=x_t$, and, if party 2 is elected, it sets $l_t=y_t$. Thus x_t^e and y_t^e are the rational expectations formed before elections of party 1's and 2's policies.

Elections are held at the beginning of the period, say period t. At the end of period $(t-1)$ the parties announce their policies for period t, that is, x_t^a, y_t^a. If the announcements are believed by voters, $x_t^e=x_t^a$ and $y_t^e=y_t^a$; otherwise, $x_t^e=E(x_t/I_{t-1})\neq x_t^a$ and $y_t^e=E(y_t/I_{t-1})\neq y_t^a$. $E(\cdot)$ is the mathematical expectation operator and I_{t-1} denotes the information set available to the voters at the end of period $t-1$. Elections take place every period. (Extensions on this point are straightforward.)

The electoral outcomes are uncertain; for any given x_t^e and y_t^e there is an associated probability of party 1's victory, which is indicated by the following function

$$(3)\qquad P_t=P(x_t^e,y_t^e).$$

This function can be derived as follows. Each voter votes for the party which is expected to deliver the policy closest to his own bliss point. However, there is uncertainty about the distribution of voter's bliss points, and in particular, about the position of the bliss point of the median voter. There may also be uncertainty about the number of abstentions, because of uncertainty about the costs of voting as perceived by different voters.[3]

---

[2] Alesina and Stephen Spear, 1987, analyze the relationship between short-lived candidates and the party as a long-lived organization.

The following assumptions on the function $P_t = P(x_t^e, y_t^e)$ are adopted:

(*i*) The function $P(x_t^e, y_t^e)$ is time invariant and "common knowledge";

(*ii*) $0 < P(\cdot) < 1$ for $x_t^e \varepsilon R$; $y_t^e \varepsilon R$;

(*iii*) $P(x_t^e, y_t^e)$ is continuous and differentiable at least twice;

(*iv*)

$$\frac{\partial P(x_t^e, y_t^e)}{\partial x_t^e} \equiv P_{x_t^e} \leq (>) 0 \quad \text{if and only if } x_t^e \geq (<) y_t^e$$

$$\frac{\partial P(x_t^e, y_t^e)}{\partial y_t^e} \equiv P_{y_t^e} \leq (>) 0 \quad \text{if and only if } x_t^e \geq (<) y_t^e.$$

Assumption (*iv*) states that if one party converges toward the other, it increases its chances of election by capturing (probabilistically) "middle voters." Note that this assumption is restrictive because it implies a prohibitive barrier to the entry of a third party for any policy followed by the two existing parties. Assumption (*iii*) is more restrictive than necessary and is adopted for simplicity. In particular, the function $P(\cdot)$ could be discontinuous along the diagonal, that is, for $x_t^e = y_t^e$. All the results of this paper generalize to this case, as shown in Alesina and Alex Cukierman (1987), in which a function such as (3) is explicitly derived from the underlying distribution of preferences of the voters.

II. Electoral Competition as a One-Shot Game

Consider one electoral competition, in isolation. Two cases should be distinguished: 1) the two parties can make binding precommitments to their preelectoral platforms; 2) no precommitments are possible.

The first case has been analyzed by Wittman (1983) and Calvert (1985). With precommitments we can impose $x_t^a = x_t^e = x_t$ and $y_t^a = y_t^e = y_t$; the announced platforms are identical to the policies expected by the voters and implemented by the elected party. The Nash equilibrium of this game can then be found by solving the following problem (the time subscripts are dropped to simplify notation):

$$\text{(4)} \quad \max_x w^1 = P(x,y)[u(x)+k] + [1-P(x,y)]u(y),$$

$$\text{(5)} \quad \max_y w^2 = P(x,y)v(x) + [1-P(x,y)][v(y)+k].$$

Following Wittman (1983) and Calvert (1985) it can be shown that the equilibrium policies ($\hat{x}$ and $\hat{y}$) satisfy the following inequalities

$$\text{(6)} \qquad c > \hat{x} > \hat{y} > 0.$$

It can also be shown that the distance between $\hat{x}$ and $\hat{y}$ is inversely related to k.[4] Calvert (1985), however, stresses that the equilibrium policies are in general very close; thus even with ideologically motivated politicians one should observe a high degree of policy convergence.

---

[3] John Ledyard, 1984, shows how to derive a function analogous to (3) from the underlying uncertainty about the costs of voting which generates an uncertain turnout. An analogous representation of electoral uncertainty can be found in Wittman (1983), and Calvert (1985).

[4] A complete proof of these results is available from the author. A sufficient condition for existence and uniqueness of this equilibrium is that the function $P(x, y)$ is concave in x and convex in y.

Consider now the case in which precommitments are not available. This is the most natural assumption because the policymaker can always change his mind and the law, when in office. Thus when, say, the candidate of party 1 is elected he does not face problem (4) any more, but he can follow his most preferred policy. It is immediate to prove the following result:

PROPOSITION 1: *For any value of k (the utility of winning an election), the time-consistent equilibrium $(\bar{x}, \bar{y})$ exists, is unique and it is given by*

$$\bar{x} = c; \qquad \bar{y} = 0. \tag{7}$$

Voters anticipate the behavior of the elected party: thus $x^e = c$ and $y^e = 0$. Then, party 1 is elected with probability $P(c,0)$; for notational simplicity $P(c,0)$ will be referred to as $\bar{P}$ throughout the rest of this paper.

Proposition 1 shows that in a one-shot game without precommitments, there cannot be any policy convergence in equilibrium, regardless of the value of k. This result implies that the convergent equilibria described in the literature are "time inconsistent," or, to put it differently, that they rely upon the assumption of binding precommitments to policy announcements.[5]

III. Electoral Competition as an Infinitely Repeated Game

The infinite repetition of the game allows the two parties to reach better equilibria than the one-shot Nash with no convergence described in Proposition 1, even without binding precommitments.

Let us consider the first-best outcome. For the case of $k = 0$, the efficient frontier of the game can be obtained by solving the following problem (where $0 < \lambda < 1$)[6]

$$\max_{x,y} \lambda\left[P(x,y)u(x) + (1-P(x,y))u(y)\right] + (1-\lambda)\left[P(x,y)v(x) + (1-P(x,y))v(y)\right]. \tag{8}$$

Note that in (8) we imposed $x^e = x$ and $y^e = y$. The solution of this problem yields the following result:

PROPOSITION 2: *The efficient frontier is given by the following policies*:

$$x^* = y^* = \lambda c.$$

The parameter λ represents the weight of party 1. The relevant section of the efficient frontier is the segment in which both parties are better off than in the one-shot Nash equilibrium. This part of the frontier is characterized by the following inequalities, labeled conditions of "individual rationality"

$$1 - \sqrt{1-\bar{P}} \le \lambda \le \sqrt{\bar{P}}. \tag{9}$$

Thus, full convergence of policies is the first-best outcome. This result is due to the concavity of the utility function of the two parties: An appropriate intermediate policy followed by both parties and, therefore, obtained with certainty makes both parties better off than the expected value of the two possible outcomes of Proposition 1 weighted by the probabilities of electoral outcomes.

Naturally, the two parties are not indifferent with respect to the choice of λ, which corresponds to a particular convergent policy. The closer λ is to one, for example, the closer is the chosen policy to the bliss point of party 1 and vice versa; thus, the higher is λ the higher the welfare of party 1. The most standard way of selecting a point on the efficient frontier is the Nash-bargaining solution. To obtain this solution we consider the one-shot Nash as the "disagreement point": in fact, the one-shot Nash can be used as a threat that the parties can make against each other at the "bargaining table."[7] This Nash-bargaining solution implies a choice of λ (say λ^*). The following result characterizes λ^*:

---

[5] In this model the equilibrium of Proposition 1 remains the unique time-consistent equilibrium if this game is finitely repeated. Partial convergence in a finitely repeated game is sustainable with asymmetric information (see Alesina-Cukierman, 1987) or with multiple one-shot Nash equilibria and strategic voting; see, however, fn. 9 on strategic voting. See also Jean Pierre Benoit and Vijay Khrishna (1985) on the possibility of cooperation in finitely repeated games with perfect information.

[6] These results generalize to any value of k. If $k \neq 0$ the two parties would still solve problem (8) to find the "first-best" policies. Then they would split k, namely the benefits of being in office: for example, they would agree to share the term of office. For a formal treatment of this particular problem see Alesina-Spear, 1987.

PROPOSITION 3: *If* λ^* *corresponds to the Nash-bargaining solution, the following holds*:

(*i*) λ^* is a function of $\bar{P}$ and only if $\bar{P}$: $\lambda^*(\bar{P})$; (*ii*) $\partial\lambda^*(\bar{P})/\partial\bar{P} > 0$; (*iii*) $\lambda^*(\frac{1}{2}) = \frac{1}{2}$; (*iv*) $\lambda^*(0) = 0$, $\lambda^*(1) = 1$.

This result is quite intuitive. An increase in $\bar{P}$ makes party 1 stronger at the bargaining table. In fact if $\bar{P}$ is higher party 1 knows that if an agreement is not reached it is more likely to be elected at the "disagreement point"; thus, this party would be better off. It follows that party 1 can impose an agreement closer to its point of view and its utility is increased. Thus, an exogenous increase of the popularity of, say, party 1, implying that $\bar{P}$ becomes higher, makes this party better off in the cooperative regime. It follows that even in a cooperative regime the two parties are not indifferent to their popularity.

An additional implication of Proposition 3 is that, for given preferences of the voters, if a party becomes more radical (in the sense that its bliss point moves away from that of the other party), its bargaining power is reduced, and the Nash-bargaining solution is further away from its most preferred policy. Suppose, for example, that c increases; then, given Proposition 3 and Assumption (*iv*) on the function $P(\cdot)$, one obtains

$$\frac{\partial\lambda^*(\bar{P})}{\partial c} = \frac{\partial\lambda^*(\bar{P})}{\partial\bar{P}}\frac{\partial\bar{P}}{\partial c} < 0. \tag{10}$$

Note that the qualitative features of Proposition (3) and of (10) are more general than the Nash-bargaining solution. In fact these features derive from the fact that a change in $\bar{P}$ affects the welfare of the two parties at the disagreement point. Any solution concept used to select a point on the efficient frontier is sensitive to such a change (qualitatively) as the Nash-bargaining solution. This is the case, for example, of the solution proposed by Ehud Kalai and Meir Smorodinsky, 1975.

Finally, it should be emphasized that the convergent electoral equilibrium $x^* = y^* = \lambda^* c$ is, in general, different from the equilibrium that would emerge as the one-shot Nash equilibrium of a game between two parties maximizing popularity.

IV. Sustainability of the Efficient Frontier and Credibility

After the election, the party in office has an incentive to follow its most preferred policy, instead of the cooperative policy on the efficient frontier. If binding commitments are unavailable, the cooperative policy is dynamically inconsistent. However, this policy may become credible if reputational considerations make a deviation from it costly enough.

---

[7] The "disagreement point" $(\bar{w}_1, \bar{w}_2)$ of this problem is

$$\begin{aligned}\bar{w}^1 &= \bar{P}u(c) + (1-\bar{P})u(0)\\ &= -\tfrac{1}{2}(1-\bar{P})c^2, \quad \text{(party 1)}\\ \bar{w}^2 &= \bar{P}v(c) + (1-\bar{P})v(0)\\ &= -\tfrac{1}{2}\bar{P}c^2. \quad \text{(party 2)}\end{aligned}$$

Thus the Nash-bargaining solution can be found by solving

$$\underset{w^1, w^2}{\text{Max}}\left[w^1 + \tfrac{1}{2}(1-\bar{P})c^2\right]\left[w^2 + \tfrac{1}{2}\bar{P}c^2\right],$$

such that $w^1 + \frac{1}{2}(1-\bar{P})c^2 \geq 0$; $w^2 + \frac{1}{2}\bar{P}c^2 \geq 0$; $w^1 = w^2 + c\sqrt{-2w^2} - \frac{1}{2}c^2$.

The last constraint is the "efficient frontier." The details of the solution are available from the author.

It is assumed that the game between the two parties is played as in Friedman, 1971. The parties "announce" a policy rule before elections. If the elected party deviates from that policy, it loses reputation and it will be expected by everybody to follow in the future its most preferred and noncooperative policy, that is, its bliss point. The other party will not cooperate anymore, since cooperation is broken. The system, therefore, reverts to the one-shot Nash. This reversion constitutes the cost of a deviation. With no loss of generality we shall assume that the reversion to the one-shot Nash lasts forever.[8]

The voters are rational and informed; therefore, they will rationally expect a reversion to the noncooperative Nash equilibrium if they observe a deviation from cooperation and vote accordingly. Thus, voters are fully rational in the sense that they use all the available information (including their knowledge of the parties' strategies) to form their forecasts. However, they are not strategic, in the sense that they do not adopt retrospective strategies, for instance, to punish a cheating party.[9]

An equilibrium is credible, therefore subgame-perfect, if the incentive to deviate from it, the "temptation," is not greater than the cost of a deviation, the "enforcement." A well-known result in the repeated game literature establishes that, if the discount factor is sufficiently close to one, every individually rational point on the efficient frontier can be sustained as a subgame-perfect equilibrium. Thus, sufficiently farsighted parties could always achieve the first-best policies on the efficient frontier.

However, there are many reasons why the discount factor might be quite low in this context. For example, one party may believe that with some probability the leadership of the opponent party may change in the future and that the new leadership may not carry over the threats of the old one. This type of uncertainty has the same effect of a low discount factor. It is easy to show by computing "temptation" and "enforcement" that the equilibrium $x^* = y^* = \lambda^*(\bar{P})c$ is sustainable as a subgame-perfect equilibrium if and only if

$$(1-\lambda^*)^2 \le q(1-\bar{P}), \tag{11a}$$

$$\lambda^{*2} \le q\bar{P}. \tag{11b}$$

We will next show that Proposition 3 and (11) imply that cooperation is more easily sustainable in a balanced system, that is, in a system in which $\bar{P}$ is close to 1/2. In fact, if $\bar{P}$ is far from 1/2, one of the two parties has little bargaining power and the policy which is selected is far from its bliss point. If this party is elected, the incentive to deviate is particularly strong. Given that this incentive is "common knowledge," it follows that cooperation is less easily sustainable. As a corollary, note that if one party adopts more extreme views, for given voters' preferences cooperation becomes more difficult (this follows immediately from (10)).

The qualitative features of these results can be illustrated using the following approximation for the Nash-bargaining solution:

$$\lambda^*(\bar{P}) = \bar{P}. \tag{12}$$

By substituting (12) into (11) it follows that the policies $x^* = y^* = \lambda^* c$ are sustainable as a subgame-perfect equilibrium if and only if the following holds,

$$q \ge \bar{P}; \qquad q \ge (1-\bar{P}). \tag{13}$$

---

[8] With no qualitative changes in the results the model could be solved for an arbitrary length of reversion to noncooperation. The length of the punishment period is completely isomorphic to the discount factor. See Friedman, 1971, on these issues. Dilip Abreu, 1983, has shown that worse "punishments" than the reversion to the one-shot Nash can be credibly threatened by the players of this type of game. However, these more general punishment strategies are not considered in this paper.

[9] In order to sustain credible equilibria by means of strategic voting, one needs to assume a high degree of coordination of strategies. In fact, an atomistic voter knows that his retrospective strategy works only if he knows the strategies of all the other voters. For a discussion along those lines of models based upon strategic behavior of atomistic agents, see for instance, Kenneth Rogoff, 1987.

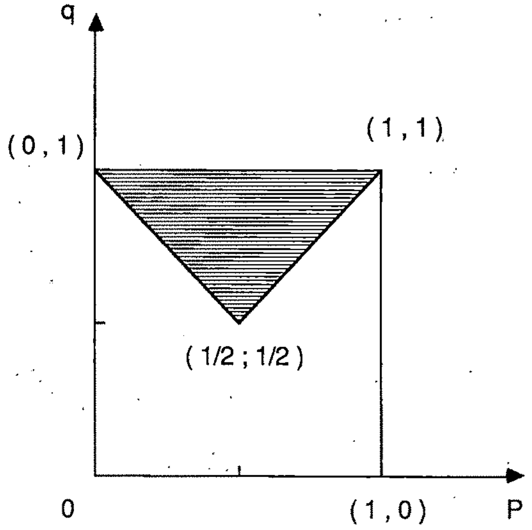


FIGURE 1

In Figure 1 the dashed area represents the locus in which the cooperative rule is sustainable. As expected, it is easiest to sustain cooperation for $\bar{P}=\frac{1}{2}$. This result is not sensitive to variations in the approximation used for λ^*.

Since we have assumed that voters do not behave strategically, the function $P(\cdot)$ does not shift in response to a deviation from cooperation. If, instead, voters used retrospective strategies (see, however, fn. 9) then, we would have a punishment arising directly from the voters. If, say, party 1 breaks the cooperative rule, there might be a shift in the function $P(\cdot)$ such that

$$(14)\quad P^{ch}(x,y)<P(x,y) \quad \text{for any } (x,y),$$

where P^{ch} identifies the probability of reelecting party 1 if this party did not follow the cooperative rule in the past. If there is a loss of popularity because of the deviation from cooperation, then, *ceteris paribus*, it is easier to sustain cooperative rules because the enforcement is stronger.[10]

V. Incomplete Cooperation

If the first-best outcome is not sustainable because q is too low, the two parties would not, in general, follow the one-shot policies (that is, their bliss points). A certain amount of convergence would still Pareto-improve upon the one-shot Nash. The best credible policies can be obtained as a solution of problem (8) under the conditions of individual rationality and the constraints of subgame perfection. The constraints of subgame perfection can be computed by imposing that for a given pair of policies $(\tilde{x},\tilde{y})$, the "temptation" of deviating from it is not greater than the "enforcement." "Temptation" and "enforcement" for the two parties can be computed following the steps described in Section IV. These constraints are

$$(15)\quad (\tilde{x}-c)^2 \le \frac{q}{1-q}\left[(1-\bar{P})c^2 - P(\tilde{x},\tilde{y})(\tilde{x}-c)^2 - [1-P(\tilde{x},\tilde{y})](\tilde{y}-c)^2\right]$$

for party 1 and

$$(16)\quad y^2 \le \frac{q}{1-q}\left[\bar{P}c^2 - P(\tilde{x},\tilde{y})\tilde{x}^2 - [1-P(\tilde{x},\tilde{y})]\tilde{y}^2\right]$$

for party 2. In the Appendix the following result is established:

PROPOSITION 4: *If and only if* $q=0$ *the only credible policies are* $x=c$ *and* $y=0$ (*the one-shot Nash*). *If* $q>0$ *there exists a nonempty set of policies* $\tilde{x}(q,c)$, $\tilde{y}(q,c)$ *which Pareto-improve upon the one-shot Nash and which satisfy the following condition*

$$c>\tilde{x}(q,c)>\tilde{y}(q,c)>0.$$

Thus, even though the one-shot Nash still remains an equilibrium of the game, there exist other, partially convergent equilibria,

---

[10] John Ferejohn, 1986, and Rogoff and Anne Sibert, 1988, consider a similar type of punishment. In their models, however, parties are not "ideological."

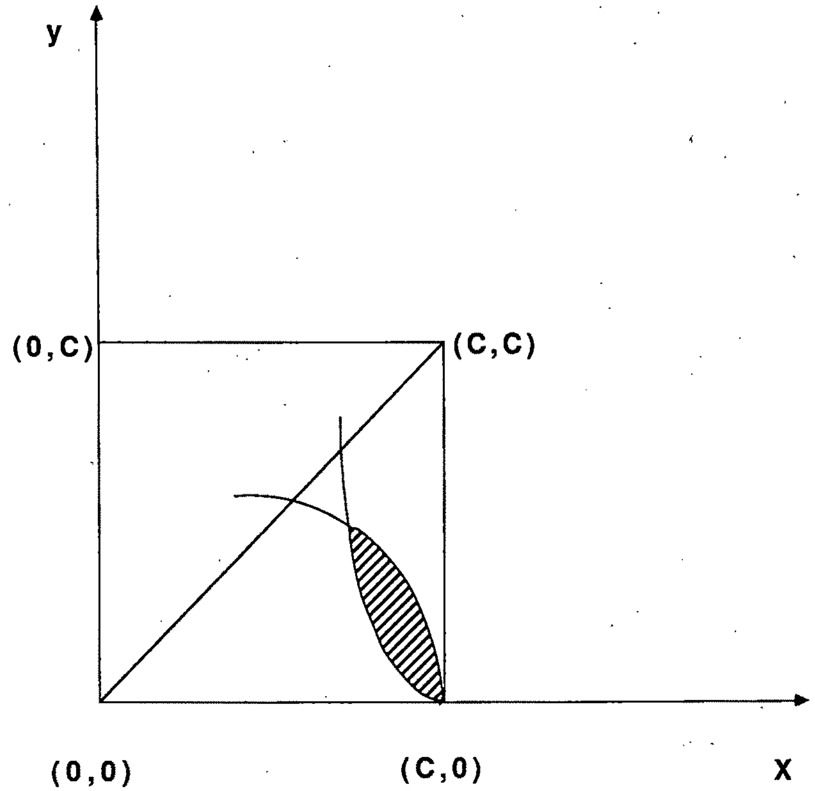


FIGURE A1

which Pareto-improve upon the one-shot Nash.

Note that the policies $(\tilde{x}, \tilde{y})$ described in Proposition 4 and the policies $(\hat{x}, \hat{y})$ of Section II, (i.e., the equilibrium policies of the one-shot game with precommitments to parties' platforms) are in general different and hardly comparable. In particular, the policies $(\hat{x}, \hat{y})$ are unrelated to the discount factor since they are the result of a one-shot game. On the contrary the set of credible policies in the infinitely repeated game $(\tilde{x}, \tilde{y})$ is crucially affected by the value of the discount factor. *Ceteris paribus*, the higher is q the larger is the set of sustainable policies.

APPENDIX: PROOF OF PROPOSITION 4

We have to solve problem (8) under the constraints of individual rationality and of subgame perfection (15) and (16). It is easy to establish that the policies $\tilde{x}$ and $\tilde{y}$ which satisfy the following conditions:

$$\text{(A1)} \qquad (i) \quad \tilde{x} \geq \tilde{y} \geq c \quad \text{or} \quad \tilde{y} > \tilde{x} \geq c.$$

$$\text{(A2)} \qquad (ii) \quad \tilde{y} \leq \tilde{x} \leq 0 \quad \text{or} \quad \tilde{x} < \tilde{y} \leq 0.$$

$$\text{(A3)} \qquad (iii) \quad \tilde{x} > c > 0 > \tilde{y}$$

cannot be a solution of the problem. These policies in fact violate at least one of the constraints of individual rationality, which can be written as

$$\text{(A4)} \quad P(x,y)(x-c)^2 + (1-P(x,y))(y-c)^2 \leq (1-\overline{P})c^2, \quad \text{(party 1)}$$

$$\text{(A5)} \quad P(x,y)x^2 + (1-P(x,y))y^2 \leq \overline{P}c^2. \quad \text{(party 2)}$$

We now show that a solution $c > \tilde{x} > \tilde{y} > 0$ always exists for any positive q. Consider the two constraints of subgame perfection. By applying the implicit function theorem it can be shown that they cross the 45° line as in Figure A1. In fact, using (15) one obtains

$$\text{(A6)} \qquad \frac{dy}{dx} = -\frac{(x-c)+\dfrac{q}{1-q}[(x-c)P(x,x)]}{\dfrac{q}{1-q}[(x-c)(1-P(x,x))]}.$$

In (A6) we used the fact that $P_x = P_y = 0$ if $x = y$. For $x < c$, (A6) implies $dy/dx < 0$. Analogous argument holds for the other constraint (16). We now show that the two curves always cross for any $q > 0$ so that an area such as the dashed area of Figure A1 always exists. Consider point $N = (c,0)$, the one-shot Nash. Consider a movement in the direction that keeps the function $P(\cdot)$ constant:

$$\text{(A7)} \qquad dP = P_x\,dx + P_y\,dy = 0.$$

Therefore

$$\text{(A8)} \qquad \frac{dy}{dx} = -\frac{P_x}{P_y} < 0.$$

Consider a movement $(c-\delta, \varepsilon)$, such that

$$\text{(A9)} \qquad \varepsilon = \frac{P_x}{P_y}\delta.$$

Define

$$\text{(A10)} \qquad \frac{P_x(0,c)}{P_y(0,c)} \equiv \beta < 0.$$

By rearranging the conditions of subgame perfection (equations (15) and (16)) at the point $(c-\delta, \varepsilon)$, and using (A9) one obtains

$$\text{(A11)} \qquad \delta \leq \frac{2\beta cq(1-\bar{P})}{1-q+\bar{P}q+\beta^2 q(1-\bar{P})},$$

$$\text{(A12)} \qquad \delta \leq \frac{2c\bar{P}q}{\beta^2(1-\bar{P}q)+\bar{P}q}.$$

Therefore, for all $q > 0$, there exists a $\delta(q)$ such that (A11) and (A12) are satisfied. It is immediately apparent that the movement $(c-\delta, \varepsilon)$, such that $\varepsilon = \beta\delta$, increases the objective function (i.e., equation (8) in the text).

REFERENCES

Abreu, Dilip, "Infinitely Repeated Games with Discounting: A General Theory and an Application to Oligopoly," unpublished doctoral dissertation, Department of Economics, Princeton University, 1983.

Alesina, Alberto, "Macroeconomic Policy in a Two-Party System as a Repeated Game," *Quarterly Journal of Economics*, August 1987, *102*, 651–78.

______, "Macroeconomics and Politics," in Stanley Fischer, ed., *NBER Macroeconomics Annual 1988*, Cambridge, MA: MIT Press, forthcoming 1988.

______ **and Cukierman, Alex,** "The Politics of Ambiguity," NBER Working Paper No. 2468, December 1987.

______ **and Sachs, Jeffrey,** "Political Parties and the Business Cycle in the United States, 1948–1984," *Journal of Money, Credit, and Banking*, February 1988, *20*, 63–82.

______ **and Spear, Stephen,** "An Overlapping Generations Model of Electoral Competition," NBER Working Paper No. 2354, August 1987.

Benoit, Jean Pierre and Khrishna, Vijay, "Finitely Repeated Games," *Econometrica*, July 1985, *53*, 905–22.

Calvert, Randall, "Robustness of the Multidimensional Voting Model, Candidate Motivations, Uncertainty and Convergence," *American Journal of Political Science*, February 1985, *39*, 69–95.

Coughlin, Peter, "Expectations about Voter Choices: A Comment," *Public Choice*, 1984, *44*, 49–59.

______ **and Nitzan, Shmuel,** "Directional and Local Electoral Equilibria with Probabilistic Voting," *Journal of Economic Theory*, May 1981, *24*, 226–39.

Downs, Anthony, *An Economic Theory of Democracy*, New York: Harper & Row, 1957.

Ferejohn, John, "Incumbent Performance and Electoral Control," *Public Choice*, August 1986, *50*, 5–26.

Friedman, James, "A Non-Cooperative Equilibrium for Supergames," *Review of Economic Studies*, January 1971, *38*, 1–12.

Hinich, Melvin, "Equilibrium in Spatial Voting: The Median Voter Result Is an Artifact," *Journal of Economic Theory*, December 1977, *16*, 208–19.

______, **Ledyard, John and Ordeshook, Peter,** "A Theory of Electoral Equilibrium: A Spatial Analysis Based on the Theory of

Games," *Journal of Politics*, February 1973, *35*, 154–93.

______, ______ and ______, "Nonvoting and the Existence of Equilibrium Under Majority Rule," *Journal of Economic Theory*, April 1972, *4*, 144–53.

______ **and Ordeshook, Peter,** "Plurality Maximization vs. Vote Maximization: A Spatial Analysis with Variable Participation," *American Political Science Review*, September 1970, *64*, 772–91.

Kalai, Ehud and Smorodinsky, Meir, "Other Solutions to Nash's Bargaining Problem," *Econometrica*, May 1975, *43*, 513–18.

Ledyard, John, "The Pure Theory of Large Two-Candidate Elections," *Public Choice*, August 1984, *48*, 7–41.

McKelvey, Richard, "Policy Related Voting and Electoral Equilibrium," *Econometrica*, September 1975, *43*, 815–63.

Rogoff, Kenneth, "Reputational Constraints on Monetary Policy," *Carnegie-Rochester Conference Series*, Spring 1987, *26*, 141–82.

______ **and Sibert, Anne,** "Equilibrium Political Business Cycles," *Review of Economic Studies*, January 1988, *55*, 1–16.

Wittman, Donald, "Candidates with Policy Preferences: A Dynamic Model," *Journal of Economic Theory*, January 1977, *15*, 86–103.

______, "Candidate Motivation: A Synthesis of Alternatives," *American Political Science Review*, March 1983, *77*, 142–57.

The Deadline Effect in Bargaining: Some Experimental Evidence

By ALVIN E. ROTH, J. KEITH MURNIGHAN, AND FRANCOISE SCHOUMAKER*

This paper reports the distribution of agreements over time in four bargaining experiments. There is a striking concentration of agreements reached in the very last seconds before the deadline. This "deadline effect" appears to be quite robust, in that the distribution of agreements over time appears to be much less sensitive to the experimental manipulations than is the distribution of the terms of agreement. Each of these experiments also exhibited a substantial frequency of disagreement.

Since last-minute agreements are widely believed to occur frequently in naturally occurring negotiations,[1] it may be helpful to state clearly just what it is that laboratory investigations have to contribute to the study of deadline phenomena. First, while there is a great deal of anecdotal information about the frequency of "eleventh-hour" agreements in naturally occurring negotiations, it has proved difficult to collect reliable data. Second, being able to study deadline phenomena in the laboratory will enable us to distinguish between alternative hypotheses in a way that the study of field data does not permit.[2] Third, while the distribution of agreements over time is one of the clearest phenomena observed in bargaining experiments to date, none of the presently available theoretical models of bargaining is able to account simultaneously for the distribution of agreements over time together with the observed patterns of agreements and substantial observed frequency of disagreements, so these results suggest clear directions for further theoretical work.

The data on the distribution of agreements over time, although analyzed here for the first time, were collected in a series of experiments in which other variables were of primary interest. In each of these experiments, in which bargaining was permitted to proceed for between nine and twelve minutes, bargaining was conducted via terminals in a computer-controlled laboratory that automatically recorded the time of each agreement. We first conducted a formal analysis of this part of the data for the experiment reported in Keith Murnighan, Alvin Roth, and Francoise Schoumaker (1988). Those re-

---

*Roth: Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260. Murnighan: Department of Business Administration, University of Illinois, 1206 S Sixth Street, Champaign, IL 61820. Schoumaker: Department de Sciences Economique, Universite de Montréal, Montréal, Quebec H3C 3J7. This work has been supported by grants from the National Science Foundation, the Office of Naval Research, and by an Alfred P. Sloan research fellowship. We have received helpful comments from Douglas Gale, Robert Forsythe, Maria Herrero, John Kennan, Roger Myerson, and Shmuel Zamir, as well as from two referees, and assistance in preparing the graphics from John Gallagher.

[1] Some representative quotations:

> Disputes over the terms of collective bargaining agreements are frequently settled only at the last minute. The last-ditch all-night parleys are as familiar to newspaper readers as they are wearing on reporters... The frequency of these photo-finishes suggests that something may be involved fundamental to the process of collective bargaining.
> [John Dunlop and James Healy, 1955, p. 57]

> The question is repeatedly asked as to why negotiations are not settled until a deadline—often at midnight or in the wee hours of the morning—even after a symbolic stopping of the clock. [Dunlop, 1984, p. 16]

> Few of us, even those least involved in bargaining activities, are unfamiliar with the 'eleventh hour' effect widely publicized in mass media accounts of collective bargaining.
> [Jeffrey Rubin and Bert Brown, 1975, p. 120]

[2] For example, labor negotiators often attribute a tendency to reach agreements just before contracts expire to the difficulty of selling any agreement to a diverse constituency if there is still time for continuing negotiations. However the deadline effect observed in our laboratory environment cannot be attributed to this, since each bargainer is bargaining strictly on his own behalf.

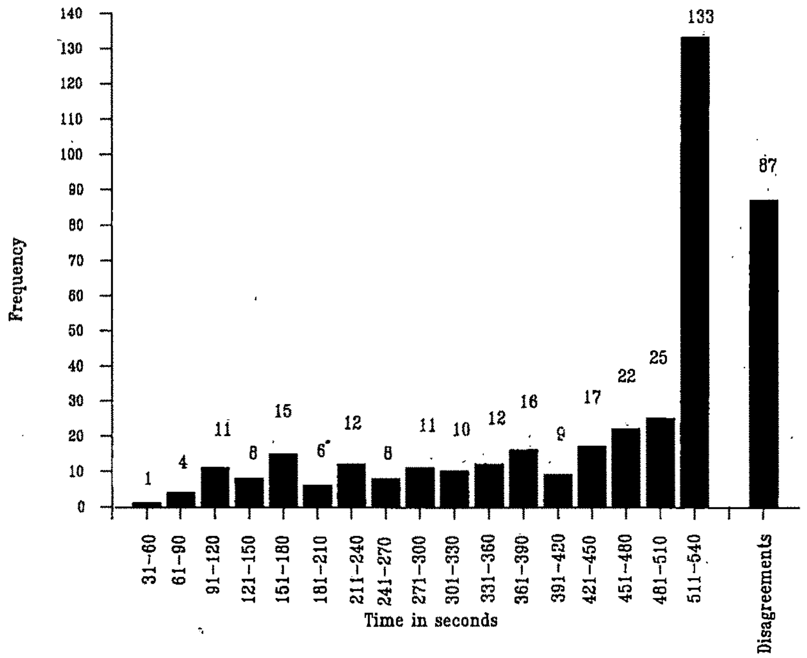


FIGURE 1A. THE FREQUENCY OF AGREEMENTS AND DISAGREEMENTS FROM THE NEW EXPERIMENT

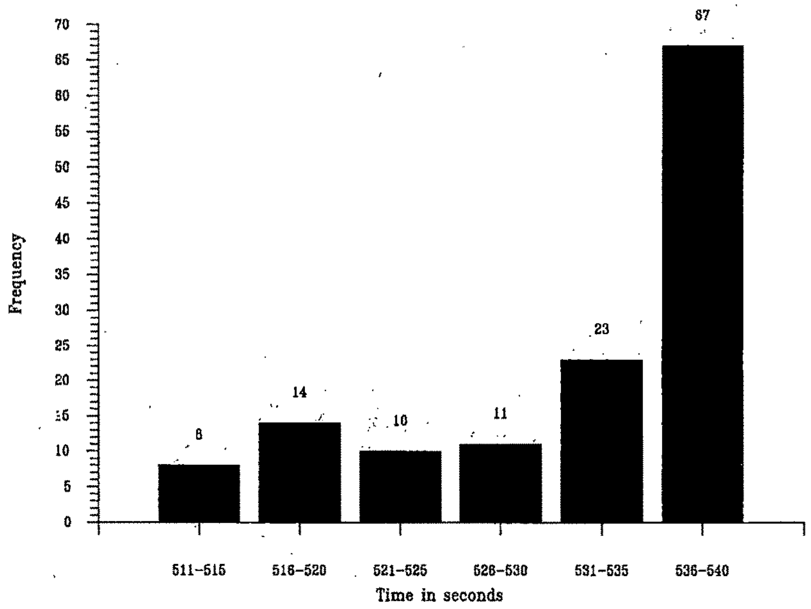


FIGURE 1B. THE FREQUENCY OF AGREEMENTS REACHED IN THE LAST 30 SECONDS OF BARGAINING IN THE NEW EXPERIMENT

sults, reported here, prompted a retrospective analysis of two earlier experiments (Roth, Michael Malouf, and Murnighan, 1981; Roth and Murnighan, 1982).[3] We also present here new experimental data which permit us to test specific hypotheses about the relationship between initial bargaining positions, bargaining outcomes, and time of agreement. Although all this data comes from laboratory experiments, there is a sense in which it is not all fully "experimental data," since the experiments were mostly designed for purposes other than to test specific hypotheses about agreement times.

The distribution of agreements over time proved to be one of the most consistent features of the results of these otherwise diverse experiments. A typical distribution, from the new experiment is shown by the histograms in Figure 1. Figure 1A graphs the number of agreements reached in each 30-second interval in the (9 minutes of) bargaining; Figure 1B breaks the last 30 seconds into 5-second intervals. Across all four experiments, 41 percent of the observed agreements were reached in the last 30 seconds of bargaining, and 53 percent of these were reached in the last 5 seconds. (Of agreements reached in the last 5 seconds, 52 percent were reached in the last second; see Table 4.)

I. The Bargaining Environment

The experiments were initially designed to test axiomatic models of bargaining of the kind proposed by John Nash (1950; see Roth, 1979), which are stated in terms of the von Neumann-Morgenstern expected utilities of the bargainers. To experimentally test these theories, an experiment must permit these utilities to be determined.

A class of games that permits this was first employed in the experiment of Roth and Malouf (1979), and is employed in three of the four experiments discussed here. In these *binary lottery games*, each agent i can win only one of two monetary prizes, a large prize λ_i or a small prize σ_i (with $\lambda_i > \sigma_i$). The players bargain over the distribution of "lottery tickets" that determine the probability of receiving the large prize: for example, an agent i who receives 40 percent of the lottery tickets has a 40 percent chance of receiving the amount λ_i and a 60 percent chance of receiving the amount σ_i. Players who do not reach agreement in the allotted time each receive σ_i. Since the information about preferences conveyed by an expected utility function is meaningfully represented only up to the arbitrary choice of origin and scale, there is no loss of generality in normalizing each agent's utility so that $u_i(\lambda_i) = 1$ and $u_i(\sigma_i) = 0$. The utility of agent i for any agreement is then precisely equal to his probability of receiving the amount λ_i, that is, equal to the percentage of lottery tickets he has received.[4] (In the specific binary lottery games discussed in this paper, σ_i always was equal to \$0.00.)

The following is typical of the procedures in each experiment.

> Each participant was seated at a visually isolated terminal of a computer system. Participants were seated at scattered terminals, and received all of their instructions and conducted all communication via the terminal.

---

[3] Data were no longer available for time of agreements in the experiment of Roth and Malouf (1979), which is the first study in the series reported here. Another related experiment (Roth and Schoumaker, 1983) is not included: negotiations in that experiment were conducted by simultaneous bids in fixed periods.

[4] We do not need to suppose in our interpretation of these experiments that the use of binary lottery games has controlled for the behavior of the experimental subjects, who may or may not be utility maximizers. Rather, the purpose of using binary lottery games is to control the predictions of the theory, specifically Nash's model of bargaining. The question of whether the various predictions of a theory like Nash's are good descriptors of behavior is independent of whether utility theory is a good description of individual choice. (That is, some of the predictions could be correct even if individuals aren't utility maximizers, and vice versa.) But since Nash's model is stated in terms of the expected utility available to the bargainers, it is necessary to control for what the utility of utility maximizers would be, to know what the predictions of the theory in any particular situation are.

Background information including a brief review of probability theory was presented first. The procedures for sending messages and proposals were then introduced. A proposal was a pair of numbers, the first being the sender's probability of receiving his prize and the second the receiver's probability. In each information condition, the terminal displayed the expected monetary value which the player would receive from any proposal he made or received. The opponent's expected value was only displayed in those conditions in which the player knew his opponent's prize. An agreement was reached whenever one of the bargainers returned a proposal identical to the one he had just received.

Bargainers could send any messages they wished, with one exception. To ensure anonymity, the monitor intercepted any messages that revealed the identity of the players. Each bargainer could make proposals at any time (regardless of what the other bargainer was doing).

The procedures varied slightly from one experiment to the next, as noted below. The time allowed for bargaining varied from nine to twelve minutes, with the time remaining coming on the screen to mark the last three minutes.

II. The Experiments

The following brief descriptions, and Table 1, provide the context needed to discuss the distribution of agreement times. Note that the various experimental manipulations accounted for significant differences in bargaining outcomes (in contrast to the uniformity we will see when we look at distributions of agreement times).

A. *The Initial Experiment of Roth and Malouf (1979)*

The most important variable in this first study manipulated the information available to the two bargainers. In the partial information condition, players knew only their own prize. In the full-information condition, players also knew one another's prize. (Theories of bargaining such as Nash's predict that the outcome of bargaining should be unaffected by the change in the information available to bargainers in the two conditions.) During the 12 minutes of bargaining, bargainers could not reveal information about their prizes in the partial information condition, and could not reveal their identity in either condition. The results were very clear: agreements were almost always equal divisions in the partial information condition, and agreements shifted significantly toward the equal expected value outcome in the full-information condition.

In this study the expected values were automatically computed for subjects to remove the effects of arithmetic ability in determining the relative success of the bargainers. The experiment discussed next was designed in part to verify that these results were not simply an artifact of making this computation available.

B. *The Experiment of Roth, Malouf, and Murnighan (1981)*

This experiment studied four games under three conditions, one in which prizes were stated in terms of an intermediate commodity, "chips." In a low-information condition, players knew only their own prizes; in an intermediate condition, players also knew their opponents' prize as valued in chips, but not money; in the full-information condition, players knew their opponents' prize in chips and in money. In the intermediate-information condition, the expected value of the prizes *in chips* was presented in exactly the same way as the expected value of the prizes in money in the previous experiment. Thus if the previous results were an artifact of providing expected values, the intermediate-information agreements would shift toward giving the player with fewer chips a higher probability of winning his prize.

The results (see Table 1) showed significant variations among outcomes. But the information about chips had no significant effect on the terms of agreement: significant shifts from equal divisions toward equal expected value agreements occurred only after

TABLE 1—A BRIEF SUMMARY OF THE FOUR EXPERIMENTS

		Differences Found Among		
Paper	Brief Description of Conditions/Studies	Terms of Agreement	Disagreement Frequencies	Agreement Times
Roth, Malouf, and Murnighan (1981)	3 information conditions 4 games	YES: A main effect for games and a games-by-information interaction	NO	NO
Roth and Murnighan (1982)	2 common knowledge conditions 4 information conditions	YES: Unimodal distributions in two info conditions; bimodal in two others; and stronger differences in common knowledge conditions	YES: One condition with higher frequencies than all others; two combined conditions different from other six	NO
Murnighan, Roth, and Schoumaker (1988)	3 studies, with different disagreement outcomes 2 games Bargainers classified as more or less risk averse	YES: (1) Disagreement outcomes, (2) the risk aversion of the bargainers, and (3) the 3 studies led to significant differences	YES: Higher disagreement outcomes led to more disagreements	YES: Study 1 agreements were faster than Study 2's or Study 3's
The New Experiment	4 games with different focal points	YES: Significant differences among the games	NO	NO

information about the monetary value of the prizes was presented.

C. *The Experiment of Roth and Murnighan (1982)*

This experiment expanded the investigation of different information conditions. The two players in each binary lottery game were assigned either a \$5 or a \$20 prize. As in the previous experiments, there was a condition in which neither bargainer was informed of the other's prize value, and a condition in which both bargainers were informed.

Two additional information conditions were included, one in which only the \$5 player knew both prizes, the other in which only the \$20 player knew both prizes. Finally, half of the conditions made this information common knowledge for both players; the other half left both parties uninformed about the other's information.

When the \$5 player knew both prizes, the terms of agreement were distributed bimodally, with the modes near equal division or equal expected value outcomes. When the \$5 player did not know both prizes, the terms of agreement approximated equal divisions, and were unimodal. More disagreements occurred when the \$5 player knew both prizes and this was not common knowledge (see Table 1).

D. *The Experiment of Murnighan, Roth, and Schoumaker (1988)*

Theoretical work (Roth, 1979; Richard Kihlstrom, Roth, and David Schmeidler, 1981; Roth and Uriel Rothblum, 1982) predicted that risk aversion should be disadvantageous in bargaining, except in situations in which the terms of the predicted agreement had a positive probability of being worse than disagreement. Three experimental studies were conducted to test these predictions. They used ternary rather than binary lottery games, in which player i has three monetary prizes, a_i, b_i, and c_i $[a_i > b_i$ and $a_i > c_i]$. The players bargain over probabilities p_1 and p_2 (with $p_2 = 1 - p_1$) such that player i receives a_i with probability p_i, and receives b_i with probability $1 - p_i$. If the players fail to reach agreement in the allotted time, then players 1 and 2 receive c_1 and c_2, respectively.

Prior to negotiating, the risk aversion of the players was assessed by having them make choices among lotteries.

In the first study, bargainers with different risk aversion were paired in two ternary lottery games, where $a_i = \$10$, $b_i = \$5$, and $c_i = \$2$ (for both bargainers) in one game, and $a_i = \$10$, $b_i = \$2$, and $c_i = \$5$ in the other. The players played each of these games twice with each of two opponents, although the fact that their two opponents were the same was disguised by the procedure. The second study changed the value of a_i, b_i, and c_i to \$10, \$5, and \$4. In the third study, one player had a high prize of \$8, and the other a high prize of \$16, with low and disagreement prizes shifting between \$4 and \$3.

The results supported the predictions: the more risk-averse bargainers did better in the high-disagreement prize game than in the low-disagreement prize game. The frequency of disagreement varied across studies and games; it was highest when the disagreement prize was high. The terms of agreement also varied significantly, with the third study (with unequal prizes for the two bargainers) exhibiting the familiar shift toward equal expected value agreements.

E. *The New Experiment*

This experiment[5] investigated different prize values in binary lottery games. The procedures were identical to those in Roth and Murnighan (1982). Each member of a bargaining pair was assigned a low prize or a high prize: prize pairs in the four conditions were (\$10 and \$15), (\$6 and \$14), (\$5 and \$20), and (\$4 and \$36). Both bargainers were informed about the value of both prizes, and this was common knowledge. So the low-prize players could argue for equal expected value divisions of lottery tickets of 60–40, 70–30, 80–20, and 90–10, respectively. Based on previous observations, we predicted that both the mean payoff to the low-prize player and the variance of the payoff distribution would increase as the prize differences increased. We also conjectured that the concentration of agreements near the deadline would be more pronounced in conditions with higher variance of agreement terms.

Unlike the Roth and Murnighan (1982) results, the data were not bimodal; distributions were approximately normal. The low-prize player's percentage of lottery tickets increased significantly as his prize decreased [$F(3,322) = 3.76$, $p < .011$], with the \$10–\$15 condition (55–45) different from the \$4–\$36 condition (61–39). The \$6–\$14 and \$5–\$20 conditions were not significantly different from each other or from the other conditions. The variance of agreement terms also differed: Bartlett's $F = 24.48$, $p < .01$, highest in the \$4–\$36 condition (std. dev. = 15.7) and lowest in the \$10–\$15 condition (std. dev. = 6.36). Disagreement frequencies were almost identical from one condition to the next: of 96 bargaining interactions in each condition, disagreements were 22, 23, 21, and 21 (23 percent overall). Thus this new data showed significant, predicted effects for prize values on the terms of agreement, but not on disagreement frequencies.

In summary, these four experiments have investigated a diverse set of experimental conditions, and both the terms of agreement and the frequency of disagreement varied significantly across these conditions (see Table 1). As we will now see, the timing of these agreements changed very little across the different conditions or experiments.

III. The Distribution of Agreements Over Time

Figure 1 presented the data from our most recent experiment. Figures 2, 3, and 4 present the data from the three earlier experiments in similar fashion. All of the figures use arbitrary time blocks of 30 and 5 seconds. Table 2 provides an additional breakdown of the data with different (but still arbitrary) time cutoffs. All four studies show a substantial number of late agreements, even as late as the last second. Although the choice of final intervals is arbitrary, there is a clear concentration of agreements near the deadline regardless of the intervals chosen to measure "nearness" to the deadline. This is evident from the cumulative distributions of agreements over time (Figures 5–8), which are all nearly vertical near the deadline.

---

[5] We hope to give a full account of other aspects of the results of this experiment in Roth, Murnighan, and Schoumaker (in preparation).

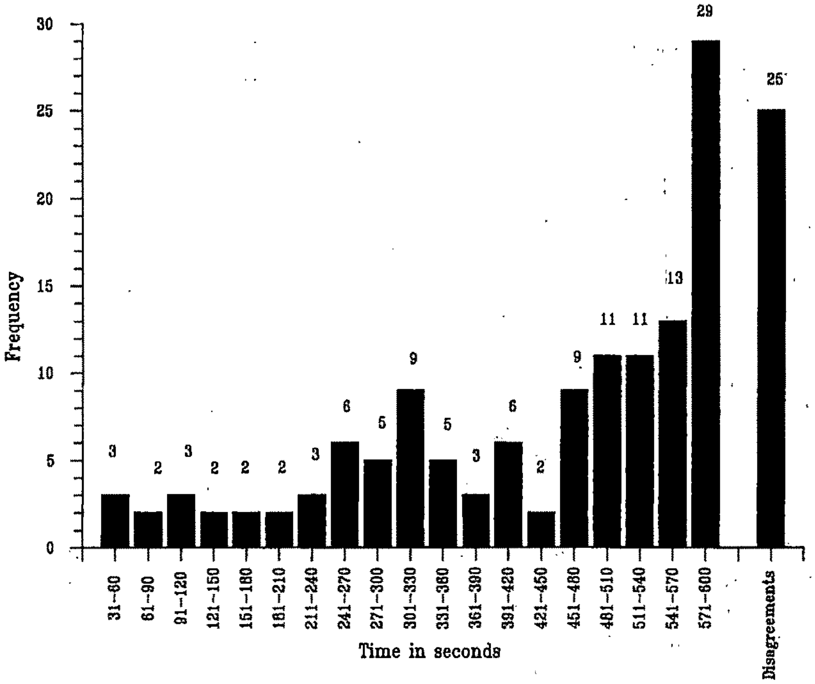


FIGURE 2A. THE FREQUENCY OF AGREEMENTS AND DISAGREEMENTS FROM ROTH, MALOUF, AND MURNIGHAN (1981)

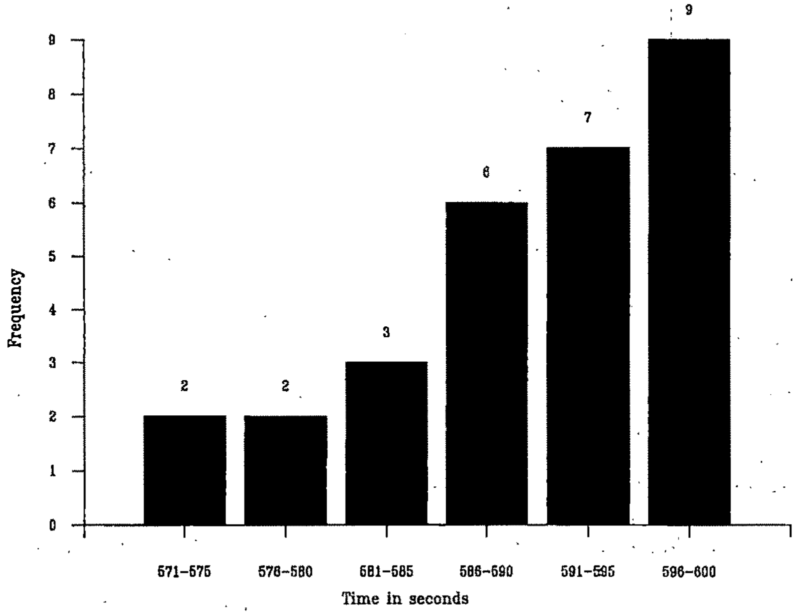


FIGURE 2B. THE FREQUENCY OF AGREEMENTS REACHED IN THE LAST 30 SECONDS OF BARGAINING IN ROTH, MALOUF, AND MURNIGHAN (1981)

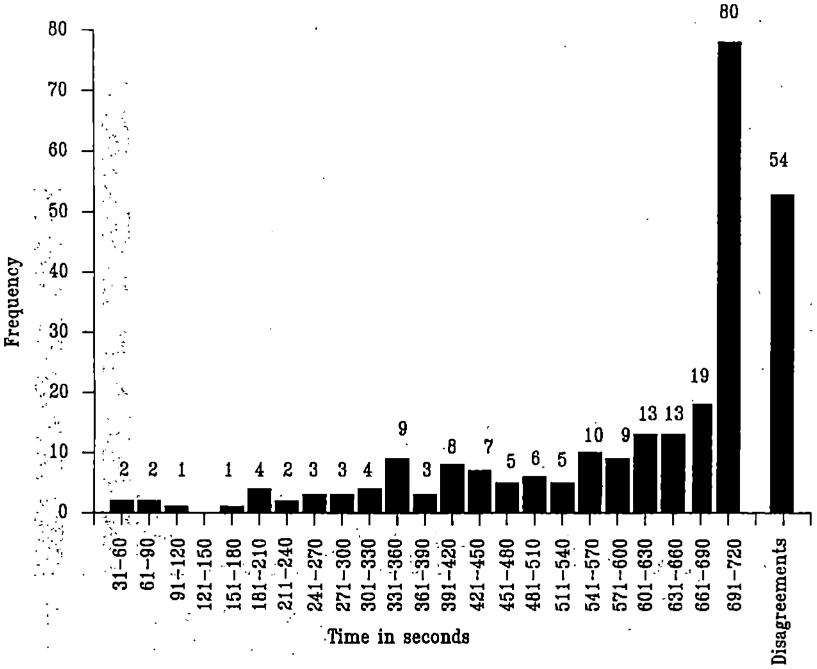


Figure 3A. The Frequency of Agreements and Disagreements from Roth and Murnighan (1982)

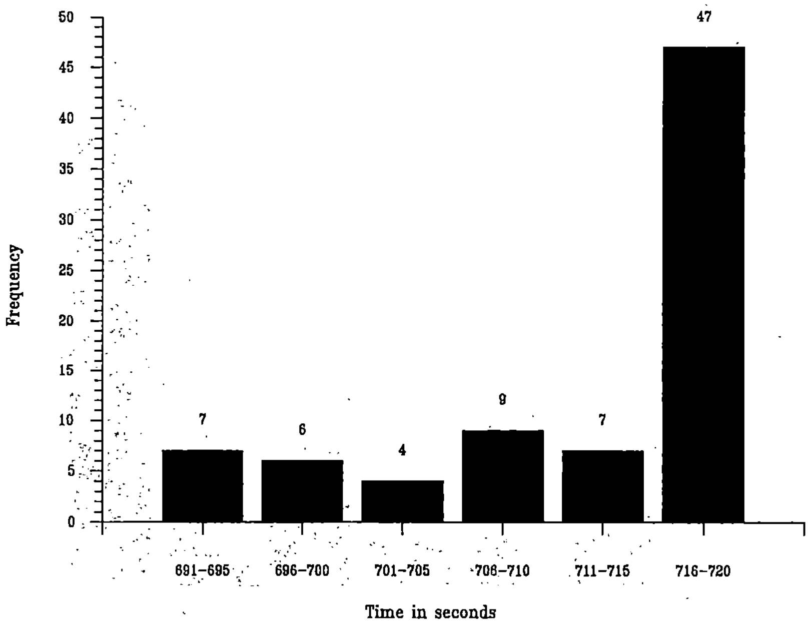


Figure 3B. The Frequency of Agreements Reached in the Last 30 Seconds of Bargaining in Roth and Murnighan (1982)

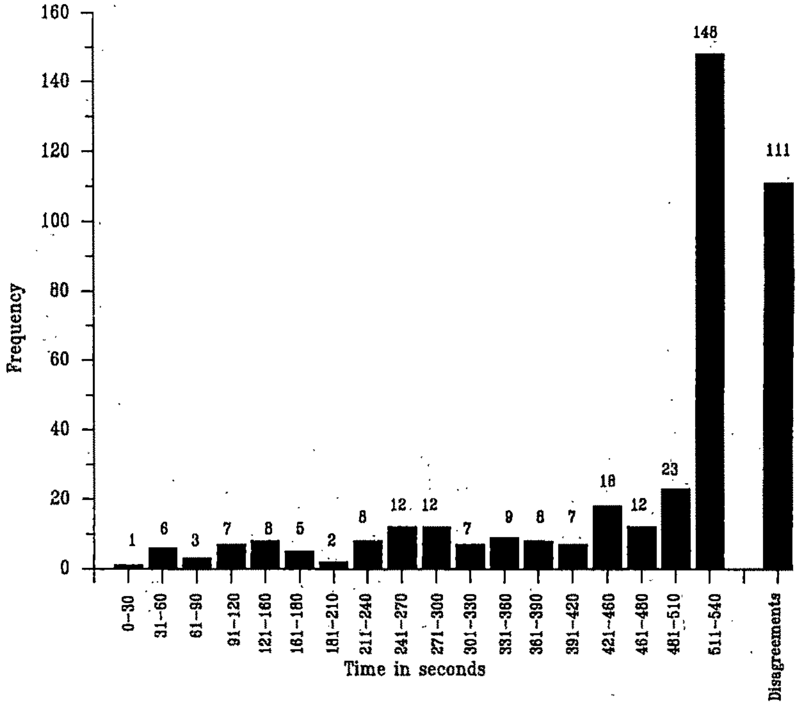


Figure 4A. The Frequency of Agreements and Disagreements from Murnighan, Roth, and Schoumaker (1988)

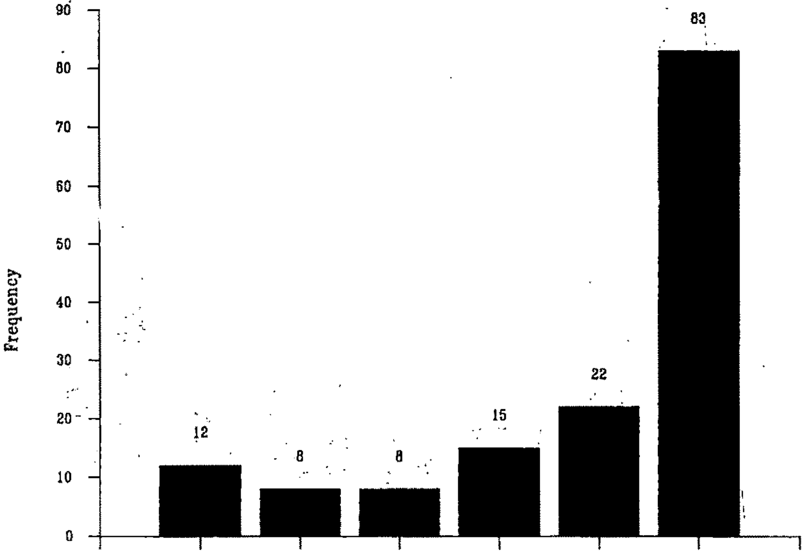


Figure 4B. The Frequency of Agreements Reached in the Last 30 Seconds of Bargaining in Murnighan, Roth, and Schoumaker (1988)

TABLE 2—THE FREQUENCY OF AGREEMENT TIMES IN FOUR EXPERIMENTS

Experiment/Condition	Total Time (seconds)	Agreements	Disagreements	Agreements in Last ___ seconds 60	45	30	15	10	5	1
Roth, Malouf, and Murnighan (1981)	600	126	25(17) (in percent)	42	37	29(19) (in percent)	23	16	9	2
Roth and Murnighan (1982)	720	209	54(21)	99	90	80(30)	63	56	46	37
Common Knowledge		93	20(18)	45	42	39(35)	30	27	18	14
Neither Knows		23	4(15)	11	11	10(37)	7	7	4	4
$20 Knows		24	6(20)	12	11	11(36)	10	8	4	4
$ 5 Knows		21	5(19)	10	10	8(32)	7	7	6	5
Both Know		25	5(17)	12	10	10(33)	6	5	4	1
Not Common Knowledge		116	34(23)	54	48	41(27)	33	29	28	23
Neither Knows		33	3(8)	16	15	12(33)	12	11	11	8
$20 Knows		20	4(17)	6	4	4(16)	4	2	1	1
$ 5 Knows		37	18(33)	23	20	17(31)	12	11	11	10
Both Know		26	9(26)	9	9	8(23)	5	5	5	4
Murnighan, Roth, and Schoumaker (1988)	540	296	111(27)	170	160	148(36)	121	108	83	30
Low-Disagreement Payoff		177	27(13)	99	94	86(42)	65	55	41	14
High-Disagreement Payoff		119	84(41)	71	66	62(31)	56	53	42	16
The New Experiment (1987)	540	325	91(22)	159	149	133(32)	103	91	67	37
10–15 Game		82	22(21)	45	44	41(39)	28	26	18	9
6–14 Game		81	23(22)	36	35	28(27)	20	18	12	7
5–20 Game		80	24(23)	40	35	33(32)	28	22	18	11
4–36 Game		82	22(21)	38	35	31(30)	27	25	19	10
Totals		956	281(23)	462	428	390(32)	310	271	205	106

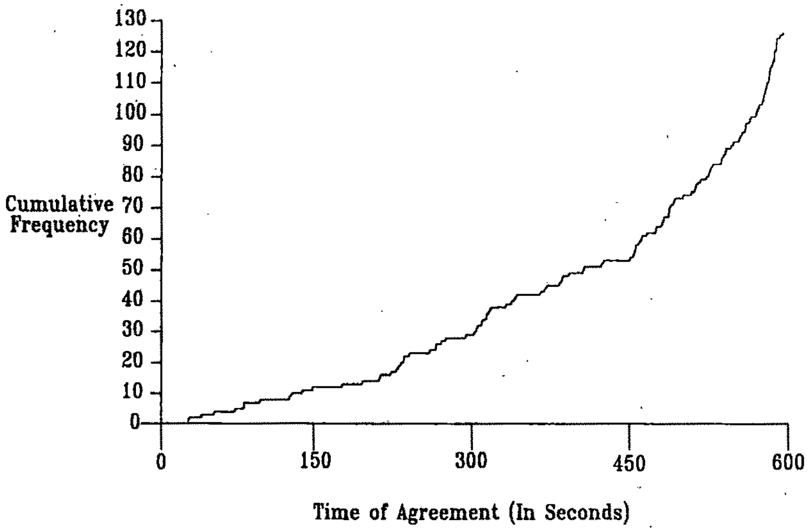


FIGURE 5. CUMULATIVE FREQUENCIES FROM ROTH, MALOUF, AND MURNIGHAN (1981)

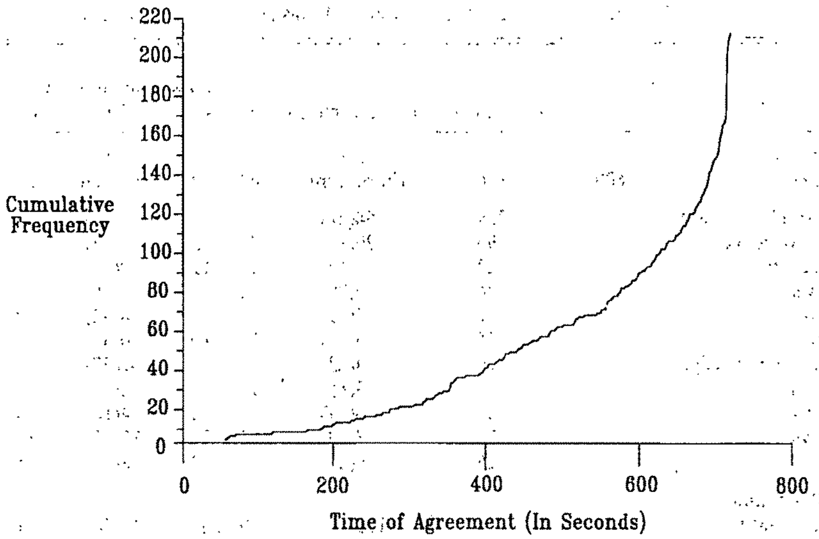


FIGURE 6. CUMULATIVE FREQUENCIES FROM ROTH AND MURNIGHAN (1982)

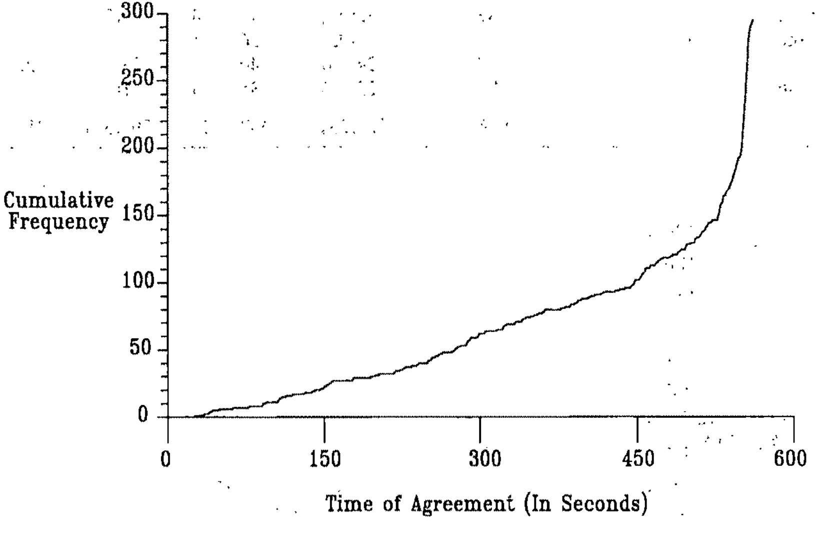


FIGURE 7. CUMULATIVE FREQUENCIES FROM MURNIGHAN, ROTH, AND SCHOUMAKER (1988)

Agreement times were compared across conditions within each experiment. Unlike the tests of agreement terms, both nonparametric and parametric tests showed no significant differences among conditions in Roth, Malouf, and Murnighan (1981), Roth and Murnighan (1982), or the new experiment. In the most stringent test, in the new experiment, neither the timing of agreements nor the variance of agreement times responded to the differences in the bargainers' prize values and the associated changes in the distribution of agreement terms.

There was a significant difference in agreement times in comparisons of the three studies in Murnighan, Roth, and Schoumaker

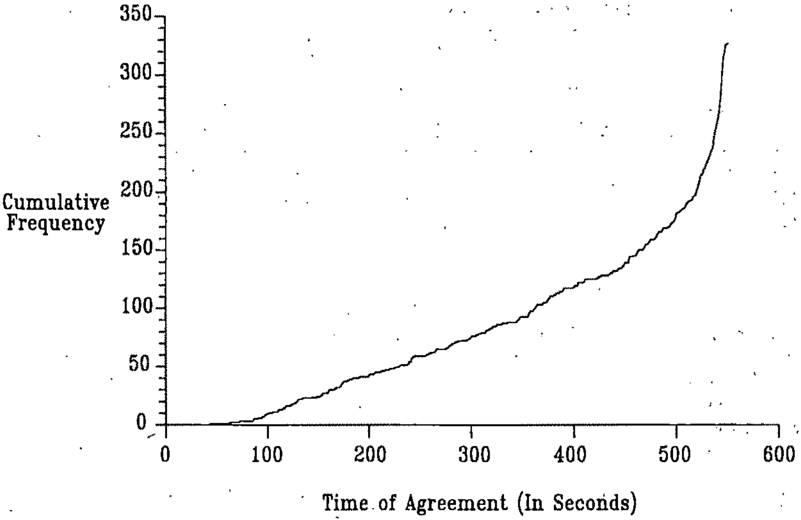


FIGURE 8. CUMULATIVE FREQUENCIES FROM THE NEW EXPERIMENT

(1988): $F(2,289)=8.32$, $p<.01$. Study 1 agreements ($\bar{X}=352$ seconds) were made more quickly than study 2 ($\bar{X}=463$) or study 3's ($\bar{X}=426$). Almost all the agreements in study 1 of this experiment were reached in the low-disagreement prize condition. In those games the players had equal prizes, and there was a preponderance of equal division agreements. The games in the other two studies were more complicated.

Analyses were also conducted within the conditions of each experiment to assess the relationship between agreement outcomes and agreement time.[6] Table 3 summarizes the correlational analyses for each of the studies. Although a few positive correlations were significant, suggesting that better outcomes were obtained by the low prize player when agreements were later, the most interesting results are in the Roth and Murnighan (1982) experiment, when the \$5 player knew both prizes and the \$20 player knew only his own. Here a strong negative correlation resulted when the information was common knowledge, and a strong positive correlation resulted when the information was not common knowledge. These correlations disappear when they are calculated for only the agreements reached in the last 30 seconds ($r(3)=-.06$ and $r(4)=-.22$, ns). Nevertheless, these findings provide the basis for speculation about the different strategies that may have been utilized in these two conditions. Overall, however, the correlations suggest little in the way of a consistent relationship between agreement outcomes and agreement times.

Table 4 provides information about the frequency of agreements in the last 30 seconds of negotiating, and how many of those agreements occurred in the last 5 or 1 second, and how many agreements in the last five seconds went to the very last second. There is clearly a concentration of agreements right before the deadline, most dramatically in Roth and Murnighan (1982), where almost half of the agreements reached in the last 30 seconds were reached in the last second.[7] These numbers are reflected in

---

[6] Analyses across conditions of the experiments also investigated the relationship between average agreement times and disagreement frequencies: the correlation was not significant.

[7] Recall that in the last three minutes of bargaining, the amount of time remaining appeared on the screen. Because of the cycle time of the software controlling the

TABLE 3—CORRELATIONS BETWEEN AGREEMENT TIME AND AGREEMENT OUTCOME FOR THE LOWER-PRIZE PLAYER (FREQUENCIES IN PARENTHESES) FOR ALL FOUR EXPERIMENTS

Roth, Malouf, and Murnighan (1981)	Game			
	1	2	3	4
Full Info	.33(12)	.07(12)	.46(12)	.11(10)
Partial Info	−.19(14)	.16(13)	−.21(10)	.33(14)
No Info	−.08(7)	−.49(8)	no variance (all 50–50's)	.83*(7)
Roth and Murnighan (1982)	Common Knowledge		Not Common Knowledge	
Neither Knows	−.15(23)		.20(33)	
$20 Knows	−.06(24)		−.16(20)	
$ 5 Knows	−.61**(21)		.49**(37)	
Both Know	−.59**(25)		−.15(26)	
Murnighan, Roth, and Schoumaker (1987)	Disagreement Prize			
	Low		High	
Study 1	−.07(32)[a]		−.28(7)[a]	
Study 2	−.06(59)[a]		−.04(32)[a]	
Study 3	.26*(84)[b]		.21(81)[b]	
The New Experiment	Game			
	10–15	6–14	5–20	4–36
	.20(82)	.03(81)	.27*(80)	−.08(82)

*$p<.05$
**$p<.01$

[a] In these conditions, both players had the same prize. The correlations were computed using the outcomes for the less risk averse of the two negotiators.

[b] In these two conditions, correlations were computed for the low-prize player. Comparable correlations for the less risk-averse players were not significant: $r(84) = -.11$ and $r(81) = -.15$, ns.

the steepness of the cumulative distributions of agreements in the final seconds of bargaining (Figures 5–8).

In summary, in each of these experiments the data exhibit a very clear concentration of agreements in the final seconds before the deadline. The distribution of times at which agreements are reached appears to be much more constant among conditions within each experiment and among experiments than are the terms of agreement (see Table 1).

IV. Possible Theoretical Models

We turn now to consider what kinds of theoretical bargaining models offer explanations for a concentration of agreements at the deadline. If we were looking for an explanation of this phenomenon in isolation, straightforward modifications of models in the literature would seem quite promising. However, the situation is more complicated when we look for an explanation of the deadline effect that can also account for the patterns of agreements and disagreements in these experiments.

---

experiments (which were conducted using a mainframe computer under time-sharing), any agreement may have been concluded a little before or after the recorded time. Similarly, the last recorded agreement was as late as 724 seconds in one "12-minute" bargaining encounter, but appears in this data as an agreement at 720 seconds.

TABLE 4—THE PERCENTAGE OF AGREEMENTS REACHED IN THE LAST 30 SECONDS AND, OF THOSE, THE PERCENTAGE REACHED IN THE LAST 5 OR 1 SECOND

	Percent Agreements in Last 30 Seconds	Percent of Last 30 Agreements in Last 5	Percent Last 30 Second Agreements in Last 1	Percent Last 5 Second Agreements in Last 1
Roth, Malouf, and Murnighan (1981)	23	31	7	22
Roth and Murnighan (1982)	38	59	46	80
Common Knowledge	42	46	36	78
Not Common Knowledge	35	66	56	85
Murnighan, Roth, and Schoumaker (1987)	50	56	20	36
Low-Disagreement Payoff	49	47	16	34
High-Disagreement Payoff	52	68	26	38
Roth, Murnighan, and Schoumaker (1987)	41	50	28	55
10–15 Game	59	43	22	50
6–14 Game	35	43	25	58
5–20 Game	41	55	33	61
4–36 Game	38	61	32	53
Totals	41	53	27	52

The family of axiomatic models of bargaining do not explicitly consider time, and do not say anything directly about the deadline effect. To consider what kinds of strategic models might be appropriate, recall the following features of the experimental environment.

To reach an agreement, a bargainer needed to return a proposal identical to the last proposal received. Since entering and sending a proposal requires a short series of key strokes, some time is needed to accept an offer, and still more time is needed to make a new proposal and have it accepted. Consequently, in the last seconds before the deadline there may be some doubt about whether enough time remains for a given transaction. Furthermore, because the computer programs that controlled these experiments had a cycle time of approximately 2 seconds, bargainers could not be completely certain how many seconds remained. Thus the exact timing of the (effective) deadline is a little uncertain. The bargainers bargained for a while with no risk that the bargaining would be exogenously terminated, but as the deadline became quite close, there was an increasing probability that the bargaining would end in the next instant, before additional proposals could be made.[8]

A. *Complete Information Models of Sequential Bargaining*

A growing body of literature concerning strategic models of bargaining explicitly models the passage of time. (Much of the recent work takes its inspiration from the infinite horizon model of Rubinstein, 1982; see Roth, 1985, for a collection of representative work.) There are two key elements to this approach. First, bargaining is modeled as being costly, in that negotiation takes time, consumption does not occur until agreement takes place, and future payoffs are discounted relative to present payoffs.

---

[8] Even apart from the precise details of sending and accepting proposals and keeping time in these experiments, it seems reasonable to postulate that in a variety of negotiating environments with a fixed deadline, as the deadline approaches, the probability rises that essentially exogenous factors will prevent an agreement from being reached.

Specifically, time is divided into periods in which bargainers alternately have the option to make offers. If an offer is accepted, bargaining ends and the bargainers receive their agreed payoffs. If an offer is declined, bargaining continues to the next period, and the value of potential agreement shrinks according to some discount factor. Disagreement results if no offer is accepted at any period. The second element is that the predicted outcome is a subgame-perfect equilibrium. Thus neither bargainer can commit himself to a contingent course of action that he would not want to carry out if the contingency should arise.

These two assumptions allow a unique subgame-perfect equilibrium to be computed. Intuitively, the shrinking value of the pie, together with the requirement of subgame perfection, allows the solution to be determined by "working backward" (compare Ingolf Stahl, 1972). The shrinking pie gives the bargainers an incentive to reach an early agreement.

Although most of the discussions in the literature speak of this shrinkage as being due to discounting, when players are expected utility maximizers it is equivalent to think of this "shrinkage" as arising from a probability that each period will be the last, so that future payoffs are discounted by the probability that the bargaining will continue to that point. That is, if a player turns down an offer in any period, there is some probability that the opportunity to continue bargaining will disappear before a counteroffer can be made, and in this case both bargainers will receive nothing (compare Binmore, Rubinstein, and Asher Wollinsky, 1986).

So the bargaining environment of our experiments can be thought of as consisting of a very large number of very short periods, each having zero probability of being the last, until we reach some period near the deadline. After that, each period has some probability of being the last, and this probability increases as the deadline approaches, until the last possible period is reached. (A finite horizon model of just this kind is considered in Joseph Harrington, 1986. See also Oliver Hart, 1986.)

Of course, in our experiments bargainers are not constrained to alternate in making offers. But the remarks we have made about these simple models, while sensitive to the rules about how offers can be made, extend, for example, to models in which each bargainer has an equal chance of being able to make the next offer at each period (see Binmore, 1987). When periods are very short, this may not be such a bad approximation to our experimental environment. Nevertheless, to develop a specific model along these lines, it would be necessary to specify exactly the rules of proposals in a way that precludes simultaneous proposals by the two bargainers,[9] but is faithful to the rules of these experiments. This could involve some subtleties which we will not explore here, since our purpose in this section is to discuss certain qualitative features of equilibrium in a model of this kind.

At a subgame-perfect equilibrium of a model of this sort, the *terms* of agreement, but not the time of agreement, are uniquely defined. (The time of agreement would be uniquely defined if the first period of bargaining had a positive probability of being the last, but we are now speaking of models in which there is no cost to delaying agreement until near the end.) At equilibrium, agreement can be reached at any period up to and including the first period that has a positive probability of being the last (i.e., at equilibrium no agreements are reached after the end of this first probabilistic period). Finally there are no disagreements at equilibrium.

While such a model does not mandate a concentration of agreements near the very end, it is consistent with it, and also consistent with some agreements being reached well before the deadline, which is what we observe. Nevertheless the predictions of such a model do not fit the data we observe. On the one hand, we observe a substantial percentage of disagreements. On the other, it is hard to argue that the observed agreements correspond to a unique equilibrium agreement, particularly in those experiments that yield a bimodal distribution of agreements.

---

[9] In order to preserve the uniqueness of the subgame-perfect equilibrium.

We can find models that predict a positive probability of disagreement at equilibrium, by looking at models of incomplete information.

B. *Incomplete Information Models of Sequential Bargaining*

In the last few years great strides have been made in exploring models of bargaining under incomplete information.[10] Whereas the complete information literature suggests that all agreements will be efficient (i.e., no disagreements will occur when joint profits are possible), various forms of inefficiency emerge as equilibrium behavior in incomplete information bargaining models. In single-period models, this inefficiency takes the form of a positive probability of disagreement at equilibrium (see, for example, Roger Myerson and Mark Satterthwaite, 1983, or Kalyan Chatterjee, 1985). In multiperiod models in which time is discounted, it takes the form of delays in the time at which agreement is reached (see, for example, Cramton, 1985; or Chatterjee, 1985). In multiperiod models in which there is some probability that each period will be the last, it also takes the form of delays, which now also imply a positive probability of disagreement, since there is a positive probability that the bargaining will terminate before agreement is reached.

To get some intuition about these models, consider a model of sequential bargaining in which (at least) one of the bargainers has some private information about something germane to the outcome of bargaining.[11] Suppose for the moment that the information is binary, and means that the informed bargainer is either in a "strong" or "weak" position, such that if the information were to become common knowledge, the informed bargainer would obtain a more favorable agreement if he is in the strong position. Even though only one bargainer is informed about whether his (own) position is strong or weak, there may still be equilibria at which a bargainer in a strong position will obtain a more favorable agreement than one in a weak position. These equilibria, which are called "separating equilibria" (since they separate strong from weak bargainers), must involve a cost to a bargainer in a strong position that he would be unwilling to pay if he were in a weak position. (Otherwise a bargainer in a weak position would simply do whatever the equilibrium would have called for him to do had he been strong.) In multiperiod models in which delay is costly, the cost to the strong bargainer of demanding the more favorable agreements is that these are reached later than agreements reached by weak bargainers.

When the costliness of delay arises from a positive probability that each period will be the last, this means that at a separating equilibrium strong bargainers face a higher percentage of disagreements than weak bargainers. Not only the existence of disagreements and delays in these models, but also the tradeoffs predicted between the frequency of disagreement and the terms of agreement, are highly evocative of the results of the experiments discussed above.[12] Nevertheless there are aspects of the data that appear difficult to reconcile with such a model. The substantial fraction of the agreements observed well before the deadline (when the probability of sudden termination is still zero) may be difficult to explain with this kind of model, particularly since the terms of early agreements do not seem to be distributed differently from late agreements.[13]

---

[10] See, for example, Chatterjee, 1985; Peter Cramton, 1985; Drew Fudenberg and Jean Tirole, 1983; Fudenberg, David Levine, and Tirole, 1985; Myerson and Satterthwaite, 1983; Myerson, 1985; Rubinstein 1985a,b; Joel Sobel and Ichiro Takahashi, 1983.

[11] In the literature this is often taken to be information about the size of the potential profits to be divided, but this is not the only kind of information that could enter the model in this way. For our present purposes, it might be useful to think of the private information as concerning the bargainer's own subjective probability distribution over the time at which a final proposal can be made.

[12] In the noncommon knowledge conditions of Roth and Murnighan (1982), the tradeoff between frequency of disagreement and the terms of agreement is virtually identical to what would be predicted at a strategic equilibrium: see the analyses presented in that paper or in Roth (1987).

V. Further Avenues of Investigation

Further experimental and theoretical work is needed to better identify the roles that deadlines play in determining the outcome of negotiations. The kinds of theoretical models discussed above suggest some directions for experiments intended to serve this purpose.[14]

Another line of investigation concerns the robustness of these results not merely to the kinds of experimental manipulations reported here, (and not only under the particular mechanics of bargaining shared by these experiments), but also to different kinds of "deadlines." In these experiments, the deadline was very sharp: after the deadline, no further bargaining could take place. In many natural bargaining situations in which eleventh-hour agreements are thought to be common (see fn. 1), the deadline is an event like the expiration of a contract. In such cases bargaining may still continue after the deadline, but at greater cost, or subject to greater uncertainties. Of course, under laboratory conditions the nature of the deadline can be freely varied, and so this too is a natural target of further experimental investigation.

In conclusion, one of the clearest phenomena to emerge from the experiments discussed here is a "deadline effect": a very high percentage of agreements are reached very close to the end, just before the deadline. This phenomenon is robust, in the sense that the distribution of agreements over time does not respond to changes in the bargaining environment nearly as much as do other features of the outcome of bargaining, such as the terms on which agreement is reached. Because deadlines are widely believed to often have similar effects in natural bargaining situations, understanding this phenomenon is likely to have practical implications about the design and conduct of negotiations, and may also shed further light on the causes of bargaining inefficiencies, which appear with substantial frequency in all the data considered here, in the form of costly disagreements. Presently available theoretical models do not appear to be able to adequately account simultaneously for the observed patterns of agreement terms, agreement times, and frequency of disagreement, although separately each of these can be "explained." It seems likely to us that many of these features of the data are related, and can neither be observed nor adequately understood in isolation. Consequently, the task of further understanding the effects of deadlines presents a challenge both to theorists and experimenters.

REFERENCES

Binmore, Ken "Perfect Equilibria in Bargaining Models," *The Economics of Bargaining*, K. Binmore and P. Dasgupta, eds., Oxford: Basil Blackwell, 1987, 77–105.

______, **Rubinstein, Ariel and Wollinsky, Asher,** "The Nash Bargaining Solution in Economic Modelling," *Rand Journal of Economics*, Summer 1986, *17*, 176–88.

---

[13] The idea is that, if bargainers in strong positions generally get better agreements by holding out longer, then late agreements should look different from early agreements. However, the absence of systematic correlations between time of agreement and terms of agreement in our data is not to surprising, in view of the fact that the experimental design makes all the observable features of the bargaining common knowledge between the bargainers in all the experiments discussed here except that of Murnighan, Roth, and Schoumaker (1987). (And in that experiment, there is no significant correlation between the time of agreement and the terms achieved by the less risk-averse bargainer; see Table 3.) Maybe if we could observe players' priors about exactly how much time was left we would see some correlation, with players who were more relaxed about the deadline doing better in last-minute agreements. This is, of course, an idea that can be investigated with an appropriately designed experiment.

[14] Of course quite different theoretical models from the ones discussed here may also suggest directions to further explore deadline phenomena. Shmuel Zamir has pointed out to us that both delays and disagreements can occur in games of timing (with complete information), in which there is some irreducible uncertainty about when a given decision to accept a standing offer will be executed.

Chatterjee, Kalyan, "Disagreement in Bargaining: Models with Incomplete Information," *Game-Theoretic Models of Bargaining*, A. E. Roth, ed., Cambridge: Cambridge University Press, 1985, 9–26.

Cramton, Peter C., "Sequential Bargaining Mechanisms," *Game-Theoretic Models of Bargaining*, A. E. Roth, ed., Cambridge: Cambridge University Press, 1985, 149–79.

Dunlop, John T., *Dispute Resolution: Negotiation and Consensus Building*, Dover, MA: Auburn House, 1984.

______ **and Healy, James J.,** *Collective Bargaining: Principles and Cases*, Homewood, IL: Richard D. Irwin, 1955.

Fudenberg, Drew and Tirole, Jean, "Sequential Bargaining with Incomplete Information," *Review of Economic Studies*, July 1983, *50*, 221–47.

______, **Levine, David and** ______, "Infinite-Horizon Models of Bargaining with One-Sided Incomplete Information," *Game-Theoretic Models of Bargaining*, A. E. Roth, ed., Cambridge: Cambridge University Press, 1985, 73–98.

Harrington, Joseph E., Jr., "A Non-Cooperative Bargaining Game with Risk Averse Players and an Uncertain Finite Horizon," *Economics Letters*, 1986, *20*, 9–13.

Hart, Oliver, "Bargaining and Strikes," Working Paper, MIT 1986.

Kihlstrom, Richard, Roth, Alvin E. and Schmeidler, David, "Risk Aversion and Solutions to Nash's Bargaining Problem," *Game Theory and Mathematical Economics*, O. Moeschlin and D. Pallaschke, eds., Amsterdam: North-Holland, 1981, 65–71.

Murnighan, J. Keith, ______ **and Schoumaker, Francoise,** "Risk Aversion in Bargaining: An Experimental Study," *Journal of Risk and Uncertainty*, 1988, *1*, 101–24.

Myerson, Roger B., "Analysis of Two Bargaining Problems with Incomplete Information," *Game-Theoretic Models of Bargaining*, A. E. Roth, ed., Cambridge: Cambridge University Press, 1985, 115–47.

______ **and Satterthwaite, Mark A.,** "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, April 1983, *29*, 265–81.

Nash, John, "The Bargaining Problem," *Econometrica*, April 1950, *18*, 155–62.

Roth, Alvin E., *Axiomatic Models of Bargaining*, Berlin: Springer, 1979.

______, **ed.,** *Game-Theoretic Models of Bargaining*, Cambridge: Cambridge University Press, 1985.

______, "Bargaining Phenomena and Bargaining Theory," *Laboratory Experimentation in Bargaining: Six Points of View*, A. E. Roth, ed., Cambridge: Cambridge University Press, 1987, 14–41.

______ **and Malouf, Michael W. K.,** "Game-Theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 1979, *86*, 574–94.

______, ______ **and Murnighan, J. Keith,** "Sociological Versus Strategic Factors in Bargaining," *Journal of Economic Behavior and Organization*, June 1981, *2*, 153–77.

______ **and Murnighan, J. Keith,** "The Role of Information in Bargaining: An Experimental Study," *Econometrica*, September 1982, *50*, 1123–42.

______ **and Schoumaker, Francoise,** "Bargaining and Coordination: Some Experiments," in preparation.

______ **and Rothblum, Uriel G.,** "Risk Aversion and Nash's Solution for Bargaining Games with Risky Outcomes, *Econometrica*, May 1982, *50*, 639–47.

______ **and Schoumaker, Francoise,** "Expectations and Reputations in Bargaining: An Experimental Study," *American Economic Review*, June 1983, *73*, 362–72.

Rubin, Jeffrey Z. and Brown, Bert R., *The Social Psychology of Bargaining and Negotiation*, New York: Academic Press, 1975.

Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model," *Econometrica*, May 1982, *50*, 97–109.

______, (1985a) "Choice of Conjectures in a Bargaining Game with Incomplete Information," *Game-Theoretic Models of Bargaining*, A. E. Roth, ed., Cambridge: Cambridge University Press, 1985, 99–114.

______, (1985b) "A Bargaining Model with Incomplete Information about Time Preferences," *Econometrica*, September 1985, *53*, 1151–72.

Sobel, Joel and Takahashi, Ichiro, "A Multi-Stage Model of Bargaining," *Review of Economic Studies*, July 1983, *50*, 411–26.

Stahl, Ingolf, *Bargaining Theory*, E.F.I., Stockholm, 1972.

A Further Test of Noncooperative Bargaining Theory: Comment

By JANET NEELIN, HUGO SONNENSCHEIN, AND MATTHEW SPIEGEL*

A great deal of attention has recently been devoted to providing noncooperative game theory foundations to bargaining problems, (compare Ingolf Stahl, 1972, and Ariel Rubinstein, 1982). This approach has produced sharp predictions about the way surplus will be divided, and has given hope that a problem long thought to be insoluble might be resolved. This paper reports the results of two experiments designed to test the theory of bargaining developed by Stahl/Rubinstein.[1] Each experiment was designed to conform precisely to their rules of the game. We were particularly interested in how their solution, which is based on backward induction, would fare when bargaining was extended over more than two periods. We were also interested in how the outcomes would compare with the alternative hypothesis that bargainers tend to split a pie 50-50.

Our work can also be seen as a response to the experiment reported by K. Binmore, A. Shaked, and J. Sutton (1985). They show that in a two-round game with alternating offers and a shrinking pie, experienced subjects chose offers consistent with the Stahl/Rubinstein theory. We duplicate their "Game B" result,[2] but demonstrate that in games with more than two rounds the Stahl/Rubinstein theory is rejected. Thus to conclude on the basis of their results that subjects behave as "gamesmen" (i.e., in a manner consistent with the predictions of backward induction), would be premature. In experiments with varying numbers of rounds, our first players consistently demanded divisions which left their opponents with shares equal to the value of the second round pie.[3] In a two-round game this behavior by definition yields demands consistent with the Stahl/Rubinstein theory. In games with more rounds it does not. The regularity of our subjects' behavior is most striking. In our experiments that went more than three rounds, the Stahl/Rubinstein division was strongly controverted. Furthermore, the divisions that were observed were remarkably similar across bargaining pairs and these divisions did not vary with experience or with the amount of money at stake.

I. The Model

In our version of the Stahl/Rubinstein model, a bargaining round consists of one party demanding a division of the "pie" which the other must accept or reject. If the

---

*Economics Department, Princeton University, Dickinson Hall, Princeton University, Princeton NJ, 08544, and Graduate School of Business Administration, Columbia University, Uris Hall, New York, NY 10027. We would like to thank the Industrial Relations Section at Princeton and the Columbia Business School for funding this project. We also thank Alvin Roth and Kenneth Binmore for insightful comments regarding the interpretation of our results. Finally, we thank Joshua Angrist Thomas Bogart, Paul Brewer, Martin Evans, Kyung Hwan Kim, Zhou Lin, John Londregan, Sheena McConnell, Phil Reny, and Jereon Swinkels for their assistance in running the experiments. Any errors are our own.

[1] Rubinstein considered the case in which there is no a priori limit on the number of demands that will be made. This case is not considered here.

[2] In their "Game A," first-time players tended to settle on an equal split. In "Game B," players who had received demands in the first game were invited to make demands. Only the opening demands in this second game were recorded. The now-experienced players acted as "gamesmen." Binmore et al. speculate that inexperienced players choose an equal split because it is an "obvious and acceptable compromise" but that players who understand the structure of the game act in a way consistent with game theory. It is natural to compare our two-round results with their "Game B" results because we allowed our subjects to play a practice round.

[3] We believe that this outcome is specific to the particular parameter values chosen. We do not claim that the Stahl/Rubinstein theory would be rejected for all choices of parameter values, nor do we claim that our choice of parameter values was the most favorable to the theory.

demand is accepted the game ends and the parties receive their shares of the pie. If the demand is rejected the game continues into another round. The power to make a demand alternates between the players. The value of the pie shrinks in each round. In our experiments, both parties receive zero payoffs if agreement is not reached in a fixed number of rounds. Suppose there are three rounds and the value of the pie in the three rounds is $5.00, $2.50, and $1.25. If rationality is common knowledge then the first player's demand is determined as follows: If play continues to the last round the first party can obtain $1.25 (less epsilon) and the second party, if that person is rational, must accept this division. Therefore, in the second round, the second party must offer at least $1.25 to the first. So if play continues to the second round the second party will get $1.25 ($2.50 − $1.25). Hence in the first round, the first party must offer $1.25 to the second, which leaves the first party with $3.75 ($5.00 − $1.25). The first party's predicted demand yields what we call the Stahl/Rubinstein split. The theory predicts that the first player will always demand the Stahl/Rubinstein split and the second player will always accept it. Hence, only one round of bargaining should ever be observed.

A. *Experiment* 1

The first experiment was conducted in two classrooms at Princeton University. The subjects were 80 junior and senior students in an intermediate micro-theory course with a calculus prerequisite. The subjects were randomly and anonymously paired; first players were assigned to a "seller" room and second players to a "buyer" room. Each subject received a set of written instructions, a sample offer sheet, and flowcharts illustrating the structure of each game. The package the subjects received is reproduced in Appendix 1. After answering questions about the instructions, we worked through an example of the four-round game on the blackboard using letters rather than numerical values since we did not want to suggest divisions to the subjects. We then ran a four-round practice game followed by two-, three-, and five-round games in which the pie started at $5.00.

TABLE 1—EXPERIMENT 1: VALUE TO BE DIVIDED IN EACH ROUND

Round Number	Two Rounds	Three Rounds	Five Rounds
1	$5.00	$5.00	$5.00
2	1.25	2.50	1.70
3	–	1.25	.58
4	–	–	.20
5	–	–	.07

Sellers made the first demand in every game. Demands were entered on a sheet which we delivered to the buyers. Buyers could accept or reject by circling the appropriate word on the sheet. The sheet was taken back and forth until the game ended. The game was structured so that buyers and sellers bargained over the share the seller would receive. Subjects knew that they would bargain with a different opponent in every game.[4]

The discount rates were chosen so that the seller's Stahl/Rubinstein demand was $3.75 in every game. By making the Stahl/Rubinstein split the same in each game we intended to give the Stahl/Rubinstein theory the best chance of being accepted. This design adds force to our rejection of the Stahl/Rubinstein prediction. We chose the value $3.75 because it is quite different from $2.50 yet it leaves the buyer with a nonnegligible amount. The values of the pie for each round of each game are shown in Table 1.[5]

---

[4] For example, seller 1 bargained first with buyer 5, then with buyer 10, and finally with buyer 15. This design was intended to minimize problems arising from adjacent buyers or sellers exchanging information about their opponents. Subjects were also asked not to discuss the game with anyone.

[5] Kenneth Binmore has pointed out to us that the "framing" of our experiment may have influenced students not to choose equal splits. Specifically, the fact that the experiment was run in the context of a micro-theory class may have predisposed students to try to act in an optimizing manner. Readers may wish to keep this caveat in mind when evaluating our rejection of the equal-split hypothesis. However, we feel that if this effect exists it strengthens our rejection of the Stahl/Rubinstein hypothesis.

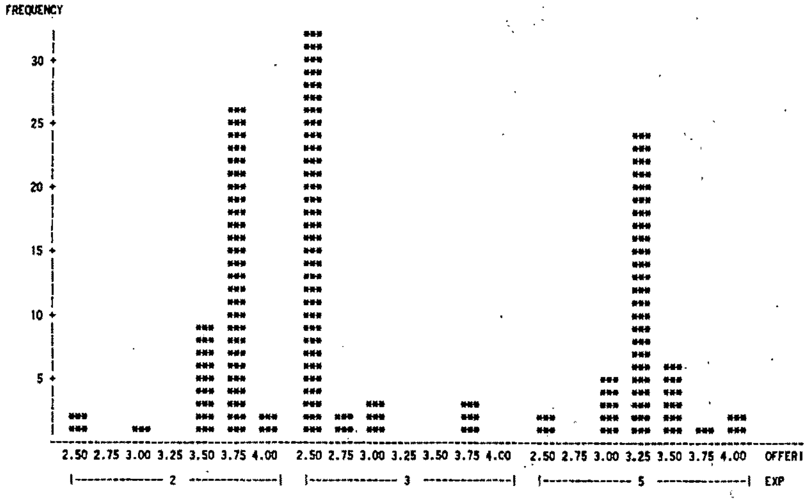


FIGURE 1. FIRST OFFERS IN THE TWO-, THREE-, AND FIVE-ROUND GAMES
FREQUENCY BAR CHART

TABLE 2—EXPERIMENT 1: VALUE OF FIRST OFFERS IN TWO-, THREE-, AND FIVE-ROUND GAMES

	Two Rounds		Three Rounds		Five Rounds	
Value	Offered	Rejected	Offered	Rejected	Offered	Rejected
< 2.50	–	–	4	–	–	–
2.50	2	–	28	–	2	–
2.51–2.99	–	–	2	–	–	–
3.00–3.29	1	–	3	–	15	1
3.30	–	–	–	–	14	1
3.31–3.49	–	–	–	–	4	–
3.50–3.74	18	2	3	2	2	–
3.75	15	3	–	–	–	–
> 3.75	4	4	–	–	3	3

The outcomes of experiment 1 are reported in Figure 1 and in Table 2. The raw data are shown in Appendix 2. Figure 1 illustrates the fact that our subjects acted very consistently: Faced with the same problem most chose similar solutions. In the two-round experiment, 15 out of 40 sellers demanded the Stahl/Rubinstein division which gave them $3.75. Thirty-three sellers made demands between $3.50 and $3.75. Sellers might be expected to ask for slightly less than $3.75 because buyers who perform the backward induction will be indifferent between accepting a demand of $3.75 and making a counteroffer of $3.75 minus epsilon. Only nine first-round demands were rejected. Of these, seven involved demands greater than or equal to $3.75. Two demands of $3.70 and $3.74 were also rejected. "Fairness" does not seem to have been at issue in these instances since neither counteroffer proposed a 50-50 split of the remaining pie. As we noted above, our two-round game replicates the "Game B" results of Binmore, Shaked, and Sutton (1985).

The Stahl/Rubinstein result is strongly controverted in our three-round experiment. Taken on its own, our three-round experi-

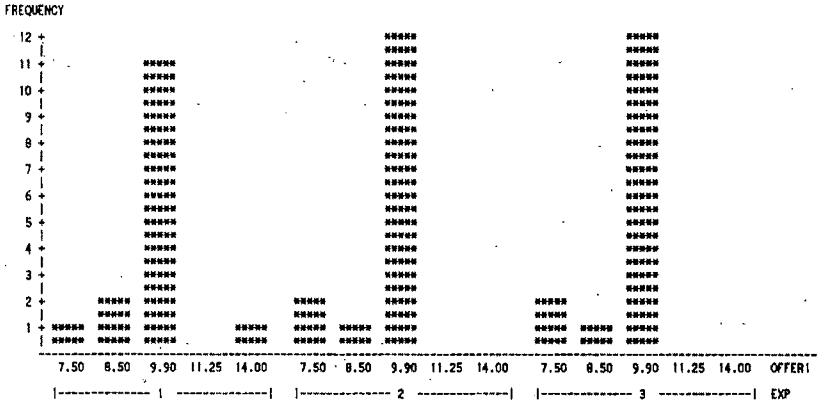


FIGURE 2. EXPERIMENT 2: FIRST OFFERS IN THE FIRST-, SECOND-, AND THIRD FIVE-ROUND GAMES FREQUENCY BAR CHART

ment could be interpreted as evidence in favor of the equal-split hypothesis. Twenty-eight out of 40 sellers demanded a 50-50 split of the ≈ 5.00 pie. Four others made demands slightly less than \$2.50 while five more shaded the other way and made demands between \$2.50 and \$3.00. Only three demanded the Stahl/Rubinstein split, and two of these demands were rejected. All of the other demands were accepted. Keep in mind the analysis of this game which follows the introduction to the model. An initial offer of \$2.50 is simply inconsistent with the hypothesis that rationality is common knowledge. Both parties should reason that if the buyer rejects a first demand, then he/she will not be able to get more than half of the remaining \$2.50 surplus. Thus, it should be common knowledge that the buyer will accept an initial demand of \$3.75, (less epsilon), and no lower demand should be observed.

Data from the five-round game are not consistent with either the Stahl/Rubinstein or the equal-split prediction. Fourteen sellers made demands of \$3.30. Altogether, 33 sellers made demands between \$3.00 and \$3.50. Only two sellers made demands between \$3.50 and \$3.75 and only two demanded \$2.50. Three demands greater than \$3.75 were made and rejected. Two demands of \$3.20 and \$3.30 were also rejected. Given that \$1.70 was the most a buyer could possibly get if bargaining continued, it is natural to consider the hypothesis that sellers act myopically: They always act as if they were in the two-round game.[6]

B. *Experiment* 2

Our second experiment was designed to test the robustness of the behavior we observed to increases in the level of bargaining experience and to increases in the amount at stake. We had thirty junior and senior intermediate micro-students play the five-round game four times: Once for practice and three times for cash.[7] We also tripled the amount of money involved. The revised flowchart for the five-round game is shown in Appendix 1.

The outcomes of experiment 2 are reported in Figure 2 and Table 3. Once again, the consistency of our subjects' behavior is

---

[6] Due to time constraints we ended the five-round game after three rounds. Only one pair was still bargaining at this point. An observant reader will notice that observation 104 in Appendix 2 is truncated.

[7] The students came from the same intermediate micro-theory course, as those in the previous experiment, but not from the same class.

TABLE 3—EXPERIMENT 2: FIRST OFFERS AND RESPONSES

Value	First Game Offered	First Game Rejected	Second Game Offered	Second Game Rejected	Third Game Offered	Third Game Rejected
Rejected						
7.50–8.00	1	–	2	–	2	–
8.75–9.00	2	–	1	–	1	–
9.50–9.85	5	1	4	1	4	–
9.90	3	2	4	1	4	–
10.00	3	–	4	–	3	–
10.50	–	–	–	–	1	1
14.00	1	1	–	–	–	–
Total Rejected		4		2		1

TABLE 4—EXPERIMENT 2: FIRST OFFERS AND RESPONSES BY INDIVIDUAL

Seller:	1	2	3	4	5	6	7
Game 1	9.90	9.50	14.00	9.90	10.00	9.50	10.00
Game 2	9.90	10.00	9.90	8.00	10.00	9.75	10.00
Game 3	9.90	10.50	9.90	9.90	10.00	9.50	10.00
Responses:[a]	111	110	011	011	111	101	111
Seller:	8	9	10	11	12	13	14	15
Game 1	9.50	10.00	9.00	9.50	9.00	9.85	9.90	8.00
Game 2	9.50	10.00	9.90	9.50	8.75	9.85	9.90	7.50
Game 3	9.50	10.00	9.90	9.50	9.00	9.85	7.50	8.00
Responses:	111	111	111	011	111	111	001	111

[a] The first digit is one if the first offer of the first game was accepted and zero otherwise. The second digit refers to the second game, etc.

striking. Almost all bargaining pairs settled in the first round, and 70 percent of the first demands lie between \$9.50 and \$10.00. There is no evidence of movement toward the Stahl/Rubinstein split on the part of the sellers, and experience does not change the opening demands. In each game, three sellers demanded less than \$9.00, 11 to 12 made demands between \$9.50 and \$10.00, and 0 or 1 demand was higher than \$10.00. By way of comparison, 24 out of 40 first demands from experiment 1's five-round game lie between \$3.20 and \$3.33. Evidently, changing the amount at stake did not significantly change the division. No demand close to the Stahl/Rubinstein demand of \$11.25 was observed.

A total of seven demands were rejected: Four in the first game, two in the second, and one in the third. Only the rejection by buyer seven of an opening bid of fourteen can be explained by noncooperative bargaining theory. The counteroffers made by the other six buyers are shown in Appendix 2.

Table 4 addresses the question of whether agents learn to become Stahl/Rubinstein gamesmen through repeated play. The opening demands are arranged by seller number. Seven sellers made the same demand in all three rounds. Four reduced their opening demand after a previous demand was rejected, which moved them further away from the Stahl/Rubinstein split. Sellers two and six raised their opening demands, but six lowered his again in the third round. Thus seller two was the only player to consistently move toward the Stahl/Rubinstein split, but her third demand was only \$10.50. In short, the hypothesis that agents will learn to be gamesmen is strongly rejected.

II. Conclusions

Neither the Stahl/Rubinstein nor the equal-split models predict the bargaining behavior observed in our six games. A convenient summary of what we observed is that in each game the sellers demanded divisions in which the buyers received the value of the second-round pie. Thus, sellers matched the best that the buyer could ever do, were players to continue to a second round. The great majority of sellers demanded this split, and their demand was accepted. The divisions that were obtained in each game were remarkably similar across bargaining pairs, and these divisions did not vary with experience or with the amount of money at stake. Indeed, the strong regularity of the behavior we observed is one of the most noteworthy aspects of our results and lends power to our rejection of both the Stahl/Rubinstein theory and the equal-split model.[8]

APPENDIX 1: EXPERIMENT 1 DATA: OFFERS AND RESPONSES IN EACH GAME, 1 = ACCEPT, 0 = REJECT

Exp	Seller	Buyer	Offer 1	Reply 1	Offer 2	Reply 2	Offer 3	Reply 3	Offer 4	Reply 4
2	31	1	3.75	1	.	.	.	.	.	.
2	32	2	3.75	1	.	.	.	.	.	.
2	33	3	3.74	1	.	.	.	.	.	.
2	34	4	3.70	1	.	.	.	.	.	.
2	35	5	3.75	0	0.75	1	.	.	.	.
2	1	6	3.75	1	.	.	.	.	.	.
2	2	7	3.90	0	0.01	0	.	.	.	.
2	3	8	3.74	1	.	.	.	.	.	.
2	4	9	3.75	1	.	.	.	.	.	.
2	5	10	3.80	0	0.65	1	.	.	.	.
2	6	11	2.50	1	.	.	.	.	.	.
2	7	12	3.50	1	.	.	.	.	.	.
2	8	13	3.70	1	.	.	.	.	.	.
2	9	14	3.75	1	.	.	.	.	.	.
2	10	15	3.74	1	.	.	.	.	.	.
2	11	16	3.85	0	0.05	0	.	.	.	.
2	12	17	3.75	1	.	.	.	.	.	.
2	13	18	3.74	0	0.25	0	.	.	.	.
2	14	19	3.50	1	.	.	.	.	.	.
2	15	20	3.75	1	.	.	.	.	.	.
2	16	21	3.50	1	.	.	.	.	.	.
2	17	22	3.75	1	.	.	.	.	.	.
2	18	23	3.75	0	0	0	.	.	.	.
2	19	24	3.50	1	.	.	.	.	.	.
2	20	25	3.00	1	.	.	.	.	.	.
2	21	26	3.75	0	0.63	1	.	.	.	.
2	22	27	3.74	1	.	.	.	.	.	.
2	23	28	3.50	1	.	.	.	.	.	.
2	24	29	3.70	0	1.00	0	.	.	.	.
2	25	30	3.60	1	.	.	.	.	.	.
2	26	31	3.75	1	.	.	.	.	.	.
2	27	32	3.74	1	.	.	.	.	.	.

---

[8] We do not claim that in general sellers will act as if they are in a two-round game. It is easy to think of games with the same structure as ours in which one might *not* expect sellers to act as if they were in a two-round game. Consider the three-round game in which the pie shrinks from $5.00 to $4.50. We would not expect the first party to demand only $.50.

Our results also do not controvert the literature which shows that in some symmetric games parties tend to settle on an equal split or on some other "focal point" (compare A. Roth and M. Malouf, 1979; or Roth and K. Murnigan, 1982), since our games have a very particular structure. For example, in our game asymmetry is introduced by giving one party the right to make the first and last demands. (We do not mean to suggest that equal splits occur only in symmetric games. See W. Güth et al., 1982, for a counterexample.)

APPENDIX 1: EXPERIMENT 1 DATA: OFFERS AND RESPONSES IN EACH GAME, 1 = ACCEPT, 0 = REJECT

Exp	Seller	Buyer	Offer 1	Reply 1	Offer 2	Reply 2	Offer 3	Reply 3	Offer 4	Reply 4
2	28	33	3.75	1	.	.	.	.	.	.
2	29	34	4.00	0	0.05	0	.	.	.	.
2	30	35	3.75	1	.	.	.	.	.	.
2	41	40	3.75	1	.	.	.	.	.	.
2	42	41	3.50	1	.	.	.	.	.	.
2	43	42	3.60	1	.	.	.	.	.	.
2	44	43	3.50	1	.	.	.	.	.	.
2	40	44	2.50	1	.	.	.	.	.	.
3	27	1	2.50	1	.	.	.	.	.	.
3	28	2	2.50	1	.	.	.	.	.	.
3	29	3	2.50	1	.	.	.	.	.	.
3	30	4	2.50	1	.	.	.	.	.	.
3	31	5	2.50	1	.	.	.	.	.	.
3	32	6	3.00	1	.	.	.	.	.	.
3	33	7	2.49	1	.	.	.	.	.	.
3	34	8	3.70	0	1.25	1	.	.	.	.
3	35	9	2.50	1	.	.	.	.	.	.
3	1	10	2.80	1	.	.	.	.	.	.
3	2	11	3.00	1	.	.	.	.	.	.
3	3	12	2.50	1	.	.	.	.	.	.
3	4	13	2.50	1	.	.	.	.	.	.
3	5	14	2.85	1	.	.	.	.	.	.
3	6	15	2.50	1	.	.	.	.	.	.
3	7	16	2.49	1	.	.	.	.	.	.
3	8	17	2.40	1	.	.	.	.	.	.
3	9	18	2.50	1	.	.	.	.	.	.
3	10	19	3.74	1	.	.	.	.	.	.
3	11	20	3.74	0	0.00	0	1.20	0	.	.
3	12	21	2.50	1	.	.	.	.	.	.
3	13	22	2.50	1	.	.	.	.	.	.
3	14	23	2.50	1	.	.	.	.	.	.
3	15	24	2.50	1	.	.	.	.	.	.
3	16	25	2.50	1	.	.	.	.	.	.
3	17	26	2.50	1	.	.	.	.	.	.
3	18	27	2.50	1	.	.	.	.	.	.
3	19	28	2.50	1	.	.	.	.	.	.
3	20	29	2.50	1	.	.	.	.	.	.
3	21	30	2.50	1	.	.	.	.	.	.
3	22	31	2.49	1	.	.	.	.	.	.
3	23	32	2.50	1	.	.	.	.	.	.
3	24	33	2.50	1	.	.	.	.	.	.
3	25	34	2.50	1	.	.	.	.	.	.
3	26	35	2.50	1	.	.	.	.	.	.
3	43	40	3.00	1	.	.	.	.	.	.
3	44	41	2.50	1	.	.	.	.	.	.
3	40	42	2.50	1	.	.	.	.	.	.
3	41	43	2.50	1	.	.	.	.	.	.
3	42	44	2.50	1	.	.	.	.	.	.
5	21	1	4.00	0	0.60	1	.	.	.	.
5	22	2	3.29	1	.	.	.	.	.	.
5	23	3	3.00	1	.	.	.	.	.	.
5	24	4	3.30	1	.	.	.	.	.	.
5	25	5	3.20	1	.	.	.	.	.	.
5	26	6	3.30	1	.	.	.	.	.	.
5	27	7	3.29	1	.	.	.	.	.	.
5	28	8	3.30	1	.	.	.	.	.	.
5	29	9	3.30	1	.	.	.	.	.	.
5	30	10	3.30	1	.	.	.	.	.	.
5	31	11	3.00	1	.	.	.	.	.	.

APPENDIX 1: CONTINUED

Exp	Seller	Buyer	Offer 1	Reply 1	Offer 2	Reply 2	Offer 3	Reply 3	Offer 4	Reply 4
5	32	12	3.40	1	·	·	·	·	·	·
5	33	13	3.29	1	·	·	·	·	·	·
5	34	14	3.29	1	·	·	·	·	·	·
5	35	15	3.30	1	·	·	·	·	·	·
5	1	16	3.40	1	·	·	·	·	·	·
5	2	17	3.53	1	·	·	·	·	·	·
5	3	18	3.29	1	·	·	·	·	·	·
5	4	19	3.25	1	·	·	·	·	·	·
5	5	20	3.40	1	·	·	·	·	·	·
5	6	21	2.50	1	·	·	·	·	·	·
5	7	22	3.29	1	·	·	·	·	·	·
5	8	23	3.20	0	0.58	1	·	·	·	·
5	9	24	3.30	0	0.01	0	0.38	·	·	·
5	10	25	4.05	0	0.65	1	·	·	·	·
5	11	26	3.87	0	0.70	·	·	·	·	·
5	12	27	3.30	1	·	·	·	·	·	·
5	13	28	3.29	1	·	·	·	·	·	·
5	14	29	3.00	1	·	·	·	·	·	·
5	15	30	3.30	1	·	·	·	·	·	·
5	16	31	3.00	1	·	·	·	·	·	·
5	17	32	3.30	1	·	·	·	·	·	·
5	18	33	3.30	1	·	·	·	·	·	·
5	19	34	3.30	1	·	·	·	·	·	·
5	20	35	3.00	1	·	·	·	·	·	·
5	44	40	3.50	1	·	·	·	·	·	·
5	40	41	2.50	1	·	·	·	·	·	·
5	41	42	3.30	1	·	·	·	·	·	·
5	42	43	3.30	1	·	·	·	·	·	·
5	43	44	3.40	1	·	·	·	·	·	·

EXPERIMENT 2 DATA: OFFERS AND RESPONSES IN EACH GAME,
1 = ACCEPT, 0 = REJECT

Exp	Seller	Buyer	Offer 1	Reply 1	Offer 2	Reply 2	Offer 3	Reply 3	Offer 4	Reply 4
0	1	1	9.90	1	·	·	·	·	·	·
0	2	2	10.00	1	·	·	·	·	·	·
0	3	3	14.00	0	4.10	1	·	·	·	·
0	4	4	7.50	1	·	·	·	·	·	·
0	5	5	7.50	1	·	·	·	·	·	·
0	6	6	9.00	1	·	·	·	·	·	·
0	7	7	9.90	1	·	·	·	·	·	·
0	8	8	9.50	1	·	·	·	·	·	·
0	9	9	10.00	1	·	·	·	·	·	·
0	10	10	7.50	1	·	·	·	·	·	·
0	11	11	9.50	1	·	·	·	·	·	·
0	12	12	9.50	1	·	·	·	·	·	·
0	13	13	9.90	0	3.10	0	1.00	0	0.0	0
0	14	14	9.90	1	·	·	·	·	·	·
0	15	15	10.00	0	2.05	0	.90	0	.30	1
1	5	1	9.90	1	·	·	·	·	·	·
1	6	2	9.50	1	·	·	·	·	·	·
1	7	3	14.00	0	3.00	1	·	·	·	·
1	8	4	9.90	0	1.75	0	1.14	0	0.30	1
1	9	5	10.00	1	·	·	·	·	·	·
1	10	6	9.50	1	·	·	·	·	·	·
1	11	7	10.00	1	·	·	·	·	·	·

EXPERIMENT 2: CONTINUED

Exp	Seller	Buyer	Offer 1	Reply 1	Offer 2	Reply 2	Offer 3	Reply 3	Offer 4	Reply 4
1	12	8	9.50	1	.	.	.	.	.	.
1	13	9	10.00	1	.	.	.	.	.	.
1	14	10	9.00	1	.	.	.	.	.	.
1	15	11	9.50	0	2.55	1	.	.	.	.
1	1	12	9.00	1	.	.	.	.	.	.
1	2	13	9.85	1	.	.	.	.	.	.
1	3	14	9.90	0	2.50	1	.	.	.	.
1	4	15	8.00	1	.	.	.	.	.	.
2	10	1	9.90	1	.	.	.	.	.	.
2	11	2	10.00	1	.	.	.	.	.	.
2	12	3	9.90	1	.	.	.	.	.	.
2	13	4	8.00	1	.	.	.	.	.	.
2	14	5	10.00	1	.	.	.	.	.	.
2	15	6	9.75	0	2.55	1	.	.	.	.
2	1	7	10.00	1	.	.	.	.	.	.
2	2	8	9.50	1	.	.	.	.	.	.
2	3	9	10.00	1	.	.	.	.	.	.
2	4	10	9.90	1	.	.	.	.	.	.
2	5	11	9.50	1	.	.	.	.	.	.
2	6	12	8.75	1	.	.	.	.	.	.
2	7	13	9.85	1	.	.	.	.	.	.
2	8	14	9.90	0	2.55	1	.	.	.	.
2	9	15	7.50	1	.	.	.	.	.	.
3	2	1	9.90	1	.	.	.	.	.	.
3	3	2	10.50	0	1.00	0	1.50	0	0.2	0
3	4	3	9.90	1	.	.	.	.	.	.
3	5	4	9.90	1	.	.	.	.	.	.
3	6	5	10.00	1	.	.	.	.	.	.
3	7	6	9.50	1	.	.	.	.	.	.
3	8	7	10.00	1	.	.	.	.	.	.
3	9	8	9.50	1	.	.	.	.	.	.
3	10	9	10.00	1	.	.	.	.	.	.
3	11	10	9.90	1	.	.	.	.	.	.
3	12	11	9.50	1	.	.	.	.	.	.
3	13	12	9.00	1	.	.	.	.	.	.
3	14	13	9.85	1	.	.	.	.	.	.
3	15	14	7.50	1	.	.	.	.	.	.
3	1	15	8.00	1	.	.	.	.	.	.

APPENDIX 2:
INSTRUCTIONS TO THE PARTICIPANTS

You are participating in an experiment about two-person bargaining. You have been divided into two groups called buyers and sellers. We will place the buyers and the sellers in different rooms and then randomly pair each buyer with a seller. The seller has an object that is worth up to $5.00 to the buyer in the first round.

There will be a two-round, a three-round, and a five-round experiment. You will have a different partner in each experiment. A round consists of one party's making an offer and the other party's accepting or rejecting it. In the first round of each experiment the seller will make an offer by filling in the appropriate blank on the offer form. The buyer can only respond by circling "accept" or "reject" on the form. We will take the forms back and forth between buyers and sellers. Please do not discuss the offers you are making or receiving. If the buyer accepts the offer the experiment ends. If the accepted offer was "x," the seller would get "x" and the buyer would get $5.00 minus "$x$." If the buyer rejects the offer the experiment continues.

In even-numbered rounds buyers make the offers and sellers can only accept or reject them, while in odd-numbered rounds the sellers make the offers. There is a penalty associated with continuing into more rounds, since the value of the object decreases with each round. In the two-round game, the value of the object decreases to $1.25 in the second round. Hence, if the buyer offers "y" and the seller accepts, the buyer gets $1.25 minus "$y$" and the seller gets "$y$." If the seller does not accept the offer both parties get nothing. In general, if bargaining continues into the last round and the last offer is not accepted, both parties get nothing.

Make sure you put your buyer number or seller number on the offer form.

The diagrams show the way bargaining proceeds in each experiment and the rate at which the value of the object shrinks with the number of rounds.

At the end of all the experiments, you will be paid the values you negotiated in cash.

These instructions are not designed to deceive or mislead you in any way. We are really only interested in what you do in this bargaining situation. In order to make things perfectly clear we will have one practice session with four rounds (for points, not cash) before beginning the experiments.

You will be discussing the theory this experiment is designed to test in class.

Thank you for your participation.

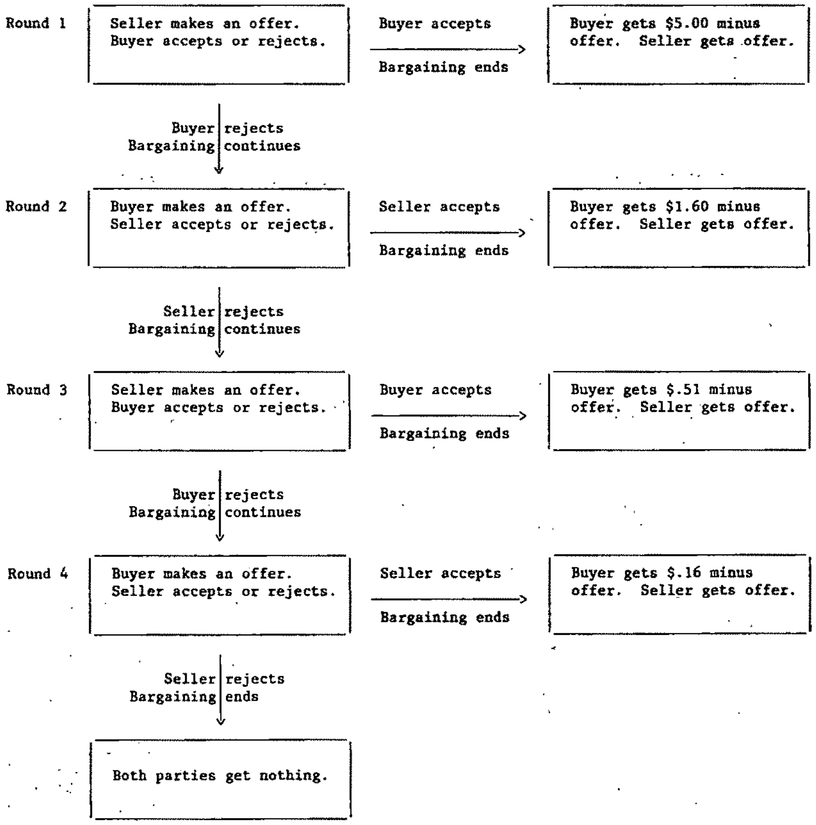

Round 1
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $5.00 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 2
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $1.60 minus offer. Seller gets offer.
Seller rejects
Bargaining continues
Round 3
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $.51 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 4
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $.16 minus offer. Seller gets offer.
Seller rejects
Bargaining ends
Both parties get nothing.

CHART 1. EXPERIMENT 1: STRUCTURE OF THE FOUR-ROUND TRIAL RUN

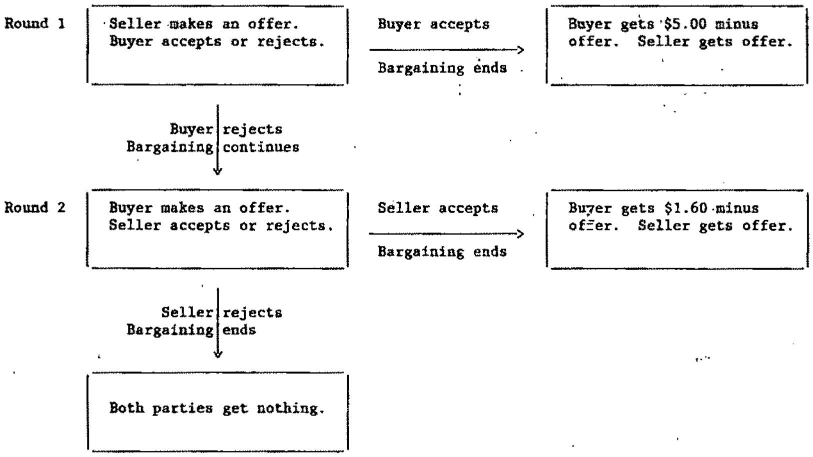

Round 1
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $5.00 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 2
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $1.60 minus offer. Seller gets offer.
Seller rejects
Bargaining ends
Both parties get nothing.

CHART 2. EXPERIMENT 1: STRUCTURE OF THE TWO-ROUND GAME

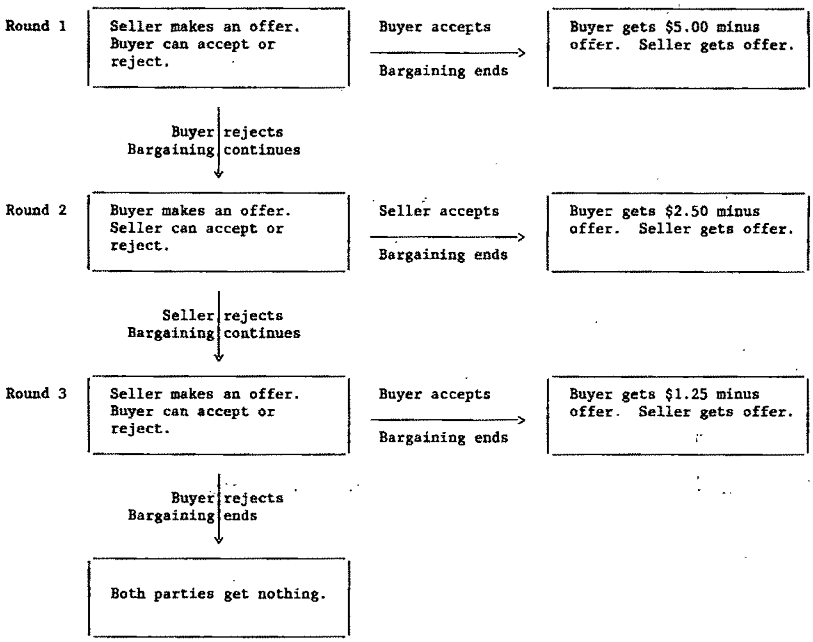

Round 1
Seller makes an offer. Buyer can accept or reject.
Buyer accepts
Bargaining ends
Buyer gets $5.00 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 2
Buyer makes an offer. Seller can accept or reject.
Seller accepts
Bargaining ends
Buyer gets $2.50 minus offer. Seller gets offer.
Seller rejects
Bargaining continues
Round 3
Seller makes an offer. Buyer can accept or reject.
Buyer accepts
Bargaining ends
Buyer gets $1.25 minus offer. Seller gets offer.
Buyer rejects
Bargaining ends
Both parties get nothing.

CHART 3. EXPERIMENT 1: STRUCTURE OF THE THREE-ROUND GAME

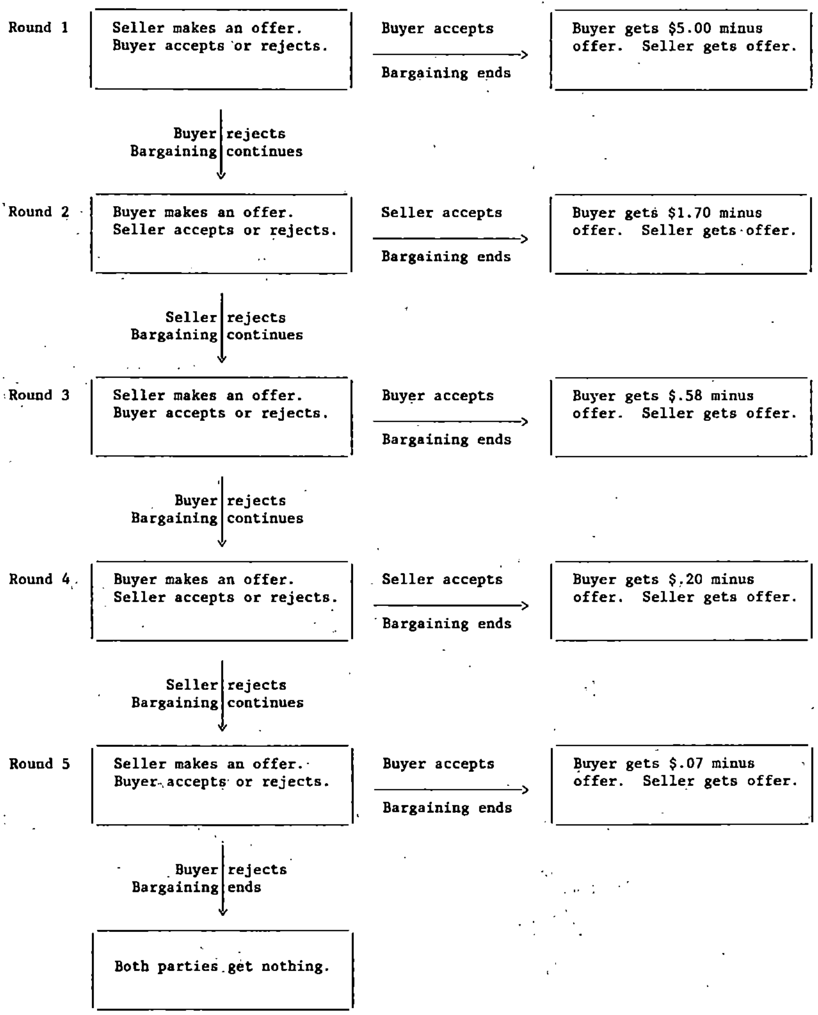

Round 1
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $5.00 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 2
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $1.70 minus offer. Seller gets offer.
Seller rejects
Bargaining continues
Round 3
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $.58 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 4
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $.20 minus offer. Seller gets offer.
Seller rejects
Bargaining continues
Round 5
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $.07 minus offer. Seller gets offer.
Buyer rejects
Bargaining ends
Both parties get nothing.

CHART 4. EXPERIMENT 1: STRUCTURE OF THE FIVE-ROUND GAME

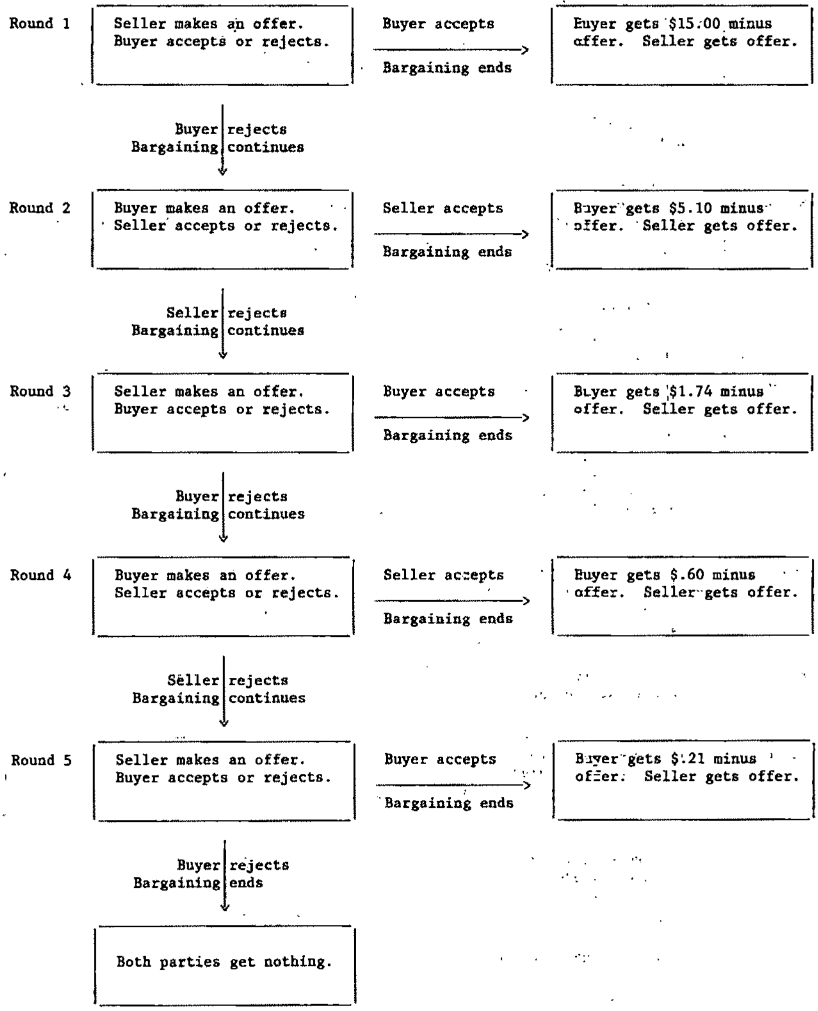

Round 1
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $15.00 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 2
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $5.10 minus offer. Seller gets offer.
Seller rejects
Bargaining continues
Round 3
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $1.74 minus offer. Seller gets offer.
Buyer rejects
Bargaining continues
Round 4
Buyer makes an offer. Seller accepts or rejects.
Seller accepts
Bargaining ends
Buyer gets $.60 minus offer. Seller gets offer.
Seller rejects
Bargaining continues
Round 5
Seller makes an offer. Buyer accepts or rejects.
Buyer accepts
Bargaining ends
Buyer gets $.21 minus offer. Seller gets offer.
Buyer rejects
Bargaining ends
Both parties get nothing.

CHART 5. EXPERIMENT 2: STRUCTURE OF THE FIVE-ROUND GAME

REFERENCES

Binmore, K., Shaked, A. and Sutton, J., "Testing Noncooperative Bargaining Theory: A Preliminary Study," *American Economic Review*, December 1985, *75*, 1178–80.

Güth, W., Schmittberger, R. and Schwarze, B., "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, December 1982, *3*, 367–88.

Roth, Alvin and Malouf, M. "Game Theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 1979, *86*, 574–94.

______ **and Murnigan, K.** "The Role of Information in Bargaining: An Experimental Study," *Econometrica*, September 1982, *50*, 1123–42.

Rubinstein, A., "Perfect Equilibrium in a Bargaining Model," *Econometrica*, January 1982, *50*, 97–109.

Stahl, I., *Bargaining Theory*, Economic Research Institute, Stockholm, 1972.

A Further Test of Noncooperative Bargaining Theory: Reply

By KEN BINMORE, AVNER SHAKED, AND JOHN SUTTON*

This note is to comment on "A Further Test of Noncooperative Bargaining Theory," by Janet Neelin, Hugo Sonnenschein, and Matthew Spiegel (1988). They conclude by rejecting both the "Stahl/Rubinstein theory" and the "equal-split model" for the bargaining games they studied.

Consider first the rejection of the "Stahl/Rubinstein theory." To begin with, I. Stahl (1972) and A. Rubinstein (1982) should be cleared of the implicit charge of naiveté. They simply proved the existence of unique subgame-perfect equilibria in certain noncooperative bargaining models. And, just as the current behavior of Kasparov and Karpov does not refute Zermelo's theorem that Chess has an optimal solution, so the experimental results of Neelin et al. (1988) are irrelevant to the theorems of Stahl and Rubinstein. Their results, however, are not irrelevant to our own experimental work, reported in K. G. Binmore, A. Shaked, and J. Sutton (1985). We studied a very simple two-stage bargaining game and observed a marked tendency for equal-split proposals to be made by inexperienced subjects, but a much stronger tendency for subjects who were slightly more experienced to make subgame-perfect proposals. Although we were much more naive then than we are now about the behavior to be expected from experimental subjects, we were very guarded in the conclusions we drew from these data. In the working paper, Binmore et al. (1984), in which our results were reported in full, we went so far as to deny explicitly that subgame-perfect equilibrium behavior should be expected in general. The reason is very simple. Even knowing the theory and being provided with the parameters of the game in a user-friendly graphical form on a microcomputer display, as in our experiment, we, ourselves, had to think quite hard to come up with the "proper" proposals in multistage versions of the Stahl/Rubinstein bargaining model. One has to find a suitable rule of thumb which relates the form in which the problem is presented to the abstract theory. Without extended opportunities for learning and ample incentives to make the necessary time-and-attention costs to the subjects worth bearing, it therefore seems unreasonable to suppose that game theory is likely to be very useful as a precise predictor of the behavior of Princeton undergraduates in this context: even if they knew game theory. This is *not* an admission that game theory is useless in economics. Game theorists defend the use of their trade in positive economics by arguing that, in the *real* economic world, there *will* often be the necessary incentives and opportunities for learning to take place. But here we are concerned with *laboratory* behavior.

But, if we were not trying to prove that game theory is a sharp predictor for laboratory bargaining behavior, what were we trying to do? Economists do not generally seem to be aware that there is large psychological literature on these matters. And, although one may have reservations about the general usefulness of psychological theory in explaining economic behavior, presumably it will be admitted that psychologists know how to run laboratory experiments. One therefore has to take their claims that "equity theory" provides a good explanation of a great deal of laboratory behavior seriously. (See, for example, Eiser, 1980; and Adams and Freedman, 1976.) In particular, Reinhard Selten (1978), the inventor of subgame-perfect equilibrium, is convinced by these claims and has his own laboratory data to offer in support of his views. It was in

---

*London School of Economics, Houghton Street, London WC2A 2AE, England.

response to a particularly striking experiment by members of his school, namely, W. Güth, R. Schmittberger, and B. Schwarze (1982), that we were led to attempt some laboratory work of our own.

The point being made here is that the issue was not, and is not, the extent to which "gamesmen" are able to solve strategic problems from cold in the laboratory. The issue is whether "gamesmen" exist at all. That is to say, whether subjects sometimes *try* to take account of strategic issues in deciding what to do. Equity theory suggests otherwise: that subjects do not act as "gamesmen" but as "fairmen." Precisely what "fair" means in any given context is a question often open to more than one interpretation. Moreover, there is the more general theory of "focal points" as described, for example, by Roth et al. (1981) to be contended with. But, although such explanatory devices are too hydra-headed to properly deserve being called "theories," the laboratory data gathered by their adherents have been replicated too often to be brushed aside as irrelevant. We, ourselves, have observed behavior that certainly looks "fair," not only from the inexperienced subjects in the study mentioned earlier, but also from subjects in later studies (Binmore et al., 1985, 1987), even though we have tried hard not to offer cues that might precipitate such behavior. This extends even to the behavior reported by E. Hoffman and M. Spitzer (1982), who observed pairs of subjects agreeing to split $14 equally, although one bargainer could guarantee himself $12 simply by withholding agreement.

Our contention in Binmore et al. (1982, 1985) was not that this behavior does not exist or is unimportant. Our contention is that it can be displaced by what we would regard as more thoughtful behavior, *given a suitable environment*. Superficially, the Neelin et al. (1988) results would seem supportive of this conclusion. But the reason that no follower of equity theory will regard their results as relevant in this context, is the manner in which their experiment was "framed." The subjects were students in an intermediate microeconomics class, who were told that they were to test a theory to be explained later. They therefore "knew" a priori that optimizing behavior was expected of them. And the evidence that subjects seek to behave "as expected" is overwhelming. For this reason, we went to great trouble (and much expense) in recruiting subjects and in designing the experimental environment, to offer as few cues as possible as to what behavior was "expected." In particular, in contrast to many studies, our own connection with the experiment was unknown to the students. Moreover, contact with our assistants and with each other was minimized by placing the subjects in separate booths and requiring them to communicate through computing equipment. Without similar precautions, even the two-stage bargaining game results of Neelin et al. are not properly comparable with ours, especially since these were preceded by a (four-stage) practice game whose results were not reported. One might remark that the experiments of W. Güth and R. Tietz (1987) are also not properly comparable with ours. (Unlike Neelin et al., they find that economics students tend to split equally in their two-stage bargaining games.)

The results of Neelin et al. are therefore perhaps best seen, not as a contribution to the debate described above, but to a theory of "bounded rationality" in simple bargaining games. The consistency with which the two-stage optimizing strategy was thoughtlessly carried over to the three- and five-stage games is certainly not a result it would have been easy to anticipate.

REFERENCES

Adams, J. and Freedman, S., "Equity Theory Revisited, Comments and Annotated Bibliography," in Berkowitz and Walster, eds., *Advances in Experimental Social Psychology*, Vol. 9, New York: Academic Press, 1976, 43–90.

Binmore, K., Shaked, A. and Sutton J., "Fairness or Gamesmanship in Bargaining: An Experimental Study," ICERD Economics Discussion Paper, London School of Economics, 1984.

______, ______ **and** ______, "Testing Noncooperative Bargaining Theory: A Preliminary Study," *American Economic Re-*

view, December 1985, *75*, 1178–80.

______, ______ **and** ______, "An Outside Option Experiment," ICERD Discussion Paper, London School of Economics, 1985 (under revision).

______, ______ **and** ______, "Do People Exploit their Bargaining Power?," forthcoming as ICERD Discussion Paper, London School of Economics, 1988.

Eiser, J., *Cognitive Social Psychology*, London: McGraw-Hill, 1980.

Güth, W., Schmittberger, R. and Schwarze, B., "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization*, December 1982, *3*, 367–88.

______ **and Tietz, R.**, "Ultimatum Bargaining for a Shrinking Cake—An Experimental Analysis," Working Paper, Frankfurt: J-W. Goethe Universitat, 1987.

Hoffman, E. and Spitzer, M., "The Coase Theorem: Some Experimental Tests," *Journal of Law and Economics*, April 1982, *25*, 73–98.

Neelin, Janet, Sonnenschein, Hugo and Spiegel, Matthew, "A Further Test of Noncooperative Bargaining Theory: Comment," *American Economic Review*, September 1988, *78*, 824–36.

Roth, A., Malouf, M. and Murnighan, J., "Sociological versus Strategic Factors in Bargaining," *Journal of Economic Behavior and Organization*, June 1981, *2*, 153–77.

Rubinstein, A., "Perfect Equilibrium in a Bargaining Model," *Econometrica*, January 1982, *50*, 97–109.

Selten, R., "The Equity Principle in Economic Behavior," in H. W. Gottinger and W. Leinfellner, eds., *Decision Theory and Social Ethics, Issues in Social Choice*, Dordrecht: Reifel Publishing, 1978, 289–301.

Stahl, I., *Bargaining Theory*, Stockholm: Stockholm School of Economics, 1972.

Preliminary Announcement of the Program

ANNUAL MEETING
THE AMERICAN ECONOMIC ASSOCIATION

New York, New York, December 27–30, 1988

Tuesday, December 27, 1988

10:00 A.M. EXECUTIVE COMMITTEE MEETING

Wednesday, December 28, 1988

8:00 A.M. ISSUES POSED BY HIGH UNEMPLOYMENT IN EUROPE
Presiding: CHARLES L. SCHULTZE, The Brookings Institution
Papers: BEN BERNANKE, Princeton University
A Reassessment of Aggregate Supply Behavior in the Interwar United States
P. R. G. LAYARD, London School of Economics, AND STEPHEN J. NICKELL, Oxford University
The Thatcher Miracle?
ROBERT J. GORDON, Northwestern University
Does Continued High Unemployment in Europe Undermine the Phillips Curve?
Discussants: BENNETT MCCALLUM, Carnegie Mellon University
J. BRADFORD DE LONG, Boston University
LAWRENCE H. SUMMERS, Harvard University

8:00 A.M. FEDERAL, STATE, AND LOCAL RELATIONS
Presiding: PAUL N. COURANT, University of Michigan
Papers: ROGER GORDON, University of Michigan
Perverse Incentives Faced by Local Governments in the People's Republic of China
ROBERT P. INMAN, University of Pennsylvania, AND DANIEL L. RUBINFELD, University of California-Berkeley
Fiscal Federalism in an Imperfect World: Lessons from the U.S.
WALLACE OATES, Resources for the Future and University of Maryland, AND ROBERT SCHWAB, University of Maryland
The Allocative and Distributive Implications of Local Fiscal Competition
HOWARD CHERNICK, Hunter College, AND ANDREW RESCHOVSKY, Tufts University
Federal Tax Reform and Fiscal Competition Among State and Local Governments
Discussants: PAUL N. COURANT, University of Michigan
HELEN F. LADD, Duke University

8:00 A.M. TAX-INDUCED DISTORTIONS IN CONSUMPTION AND LABOR SUPPLY
Presiding: DON FULLERTON, University of Virginia
Papers: PATRICIA APPS, Australian National University
Modeling Household Labor Supply and the Welfare Effects of Tax Reforms
LEONARD E. BURMAN AND B. K. ATROSTIC, U.S. Department of the Treasury
A General Equilibrium Analysis of Taxation of Fringe Benefits
CAROL ANN ROGERS, Georgetown University
Equity, Efficiency, and Time-Consistent Expenditure and Income Taxation
Discussants: MARK KILLINGSWORTH, Rutgers University
JOHN WEINBERG, Purdue University
STEVE WOODBURY, Michigan State University and Upjohn Institute

8:00 A.M. ECONOMICS OF THE SECONDARY MORTGAGE MARKETS
Presiding: LAWRENCE WHITE, Federal Home Loan Bank Board
Papers: DWIGHT M. JAFFEE, Princeton University, AND KENNETH T. ROSEN, University of California-Berkeley
Moral Hazard in the Secondary Mortgage Markets
PATRIC H. HENDERSHOTT, Ohio State University, AND ROBERT VAN ORDER, Freddie Mac
Secondary Mortgage Markets and Mortgage Pricing

EDUARDO S. SCHWARTZ AND WALTER TOROUS, University of California-Los Angeles
Valuation of Mortgage-Backed Securities
Discussants: STUART I. GREENBAUM, Northwestern University
ERIC I. HEMEL, First Boston Corporation
SUSAN WOODWARD, U.S. Department of Housing and Urban Development

8:00 A.M. SMALL BUSINESS AND ENTREPRENEURSHIP
Presiding: WILLIAM J. LAWRENCE, Pace University
Papers: BRUCE PHILLIPS, U.S. Small Business Administration
Birth and Death Rates in Small Versus Large Firms, 1976–1986
SARI SCHIENBERG, New York University, AND WILLIAM J. LAWRENCE, Pace University
Economic Determinates Impacting Small Business Longevity and Success
HELEN M. HUNTER, Bryn Mawr College
The Contrasting Financial Behavior of Small and Large Firms: Implications for Financial Stability and the Effect of Monetary Policy Since 1950
K. MARK WEAVER, University of Alabama
Key Factors to Differentiate Small Export Firms and Non-Export Firms in the United States
Discussants: CARL CREGO, Pace University
BARBARA S. MANN, U.S. Department of the Treasury
BRONWYN H. HALL, National Bureau of Economic Research
LESLIE E. PAPKE, Boston University

8:00 A.M. LESSONS FROM NUCLEAR POWER REGULATION
Presiding: PAUL JOSKOW, Massachusetts Institute of Technology
Papers: GLENN WOROCH, GTE Laboratories
Dividing a Sunk Cost: An Application of Bargaining to Power Plant Cancellations
GEOFF ROTHWELL, Stanford University
Do or Die: Operating Nuclear Reactors with Emergency Shutdown
RICHARD LESTER, Massachusetts Institute of Technology
The Effects of Industrial Organization on Learning-by-Using in Nuclear Power Plant Operation
JONATHAN FEINSTEIN, Stanford University
Safety Inspection of Commercial Nuclear Power Plants: A Policy Proposal
Discussants: MARTIN ZIMMERMAN, Ford Motor Company
LEWIS PERL, National Economic Research Associates
PAUL JOSKOW, Massachusetts Institute of Technology
LINDA COHEN, University of California-Irvine

8:00 A.M. GAMBLING AND PUBLIC POLICY
Presiding: WILLIAM EADINGTON, University of Nevada-Reno
Papers: GABRIELLE A. BRENNER AND REUVEN BRENNER, University of Montreal
A Profile of Gamblers
CHARLES CLOTFELTER AND PHILIP J. COOK, Duke University
Policy Choices for State Lotteries
REUVEN BRENNER, University of Montreal
Governments, Legislations, and Games of Chance
WILLIAM EADINGTON, University of Nevada-Reno
Balancing Economic and Social Tradeoffs: Indication of the Spread of Commercial Gambling in the 1980s
Discussants: COLIN CAMERER, University of Pennsylvania
HAROLD HOCHMAN, City University of New York
JOHN NYE, Washington University

8:00 A.M. EQUAL EMPLOYMENT OPPORTUNITY FOR WOMEN IN THE 1980s
Presiding: BARBARA BERGMANN, University of Maryland
Papers: FRANCINE BLAU, University of Illinois-Urbana-Champaign
Occupational Segregation by Gender: A Look at the 1980s
JONATHAN LEONARD, University of California-Berkeley
Women and Affirmative Action in the 1980s
JUNE O'NEILL, Baruch College, MICHAEL BRIEN, University of Chicago, AND JAMES CUNNINGHAM, U.S. Commission on Civil Rights
Discussants: RONALD EHRENBERG, Cornell University
SHELLY LUNDBERG, University of Washington

8:00 A.M. ESCALATING URBAN LAND VALUES
Presiding: BENJAMIN CHINITZ, Lincoln Institute of Land Policy
Papers: KARL CASE, Wellesley College
The Impact of Escalating Land Values on Affordable Housing
MARILYN SWARTZ LLOYD, Economic Development and Industrial Corporation
Protecting the Manufacturing Sector in the Central City
R. D. NORTON, Bryant College
Why Such Rapid Escalation in the 1980s?
Discussants: MARC WEISS, Massachusetts Institute of Technology
DAVID DOWALL, Harvard University
ROGER BOLTON, Williams College

8:00 A.M. NATIONAL SAVING AND INVESTMENT: CAUSES AND CONSEQUENCES
Presiding: JAMES GAPINSKI, Florida State University
Papers: KENNETH LEWIS, University of Delaware, RANDY NELSON, Colby College, AND LAURENCE SEIDMAN, University of Delaware
If We Raise the Saving Rate, How Long Do We Sacrifice?
EUGENE KROCH, University of Pennsylvania
Public Capital Formation and Private Saving
DENIS KESSLER, University of Paris, SERGIO PERELMAN, University of Liège, AND PIERRE PESTIEAU, Cornell University and University of Liège
Why Do Saving Rates in OECD Countries Diverge?
Discussants: JAMES GAPINSKI, Florida State University
KRISS SJOBLOM, Haverford College
JAMES BUTKIEWICZ, University of Delaware

8:00 A.M. MACROECONOMIC ADJUSTMENT IN THE EAST ASIAN NICs (Joint Session with the American Committee on Asian Economic Studies)
Presiding: JAMES RIEDEL, Johns Hopkins University
Papers: BELA BALASSA, Johns Hopkins University
Adjustment Policies in East Asia
PETER PETRI, Brandeis University
Macroeconomic Balances and Economic Structure in East Asia
NATHANIEL H. LEFF AND KAZUO SATO, Columbia University
Saving, Investment, and the Macroeconomic Opportunity Set: East Asia and Latin America
Discussants: ROMEO BAUTISTA, International Food Policy Research Institute
JAGDISH BHANDARI, International Monetary Fund
MICHAEL LEWIN, Johns Hopkins University

8:00 A.M. ETHICAL PREMISES IN ECONOMIC THOUGHT (Joint Session with the Association for Social Economics)
Presiding: HANS E. JENSEN, University of Tennessee-Knoxville
Papers: JEFFREY YOUNG, St. Lawrence University
Social Justice and Economic Thought: Adam Smith
BARRY S. CLARK, Pennsylvania State University
Social Justice and Economic Thought: John Stuart Mill
DUDLEY DILLARD, University of Maryland
Value Premises in the Economic Thought of J. M. Keynes
Discussants: JAMES HENDERSON, Valparaiso University
DONALD MOGGRIDGE, University of Toronto

8:00 A.M. PRICE STABILIZATION POLICIES IN PAST AND RECENT CHINA (Joint Session with the Chinese Economic Association in North America)
Presiding: TEH-WEI HU, University of California-Berkeley
Papers: DE-PIAO TANG, University of Missouri
Ending Hyperinflation: Lessons from the Chinese Experience in the Early 1950s
WILLIAM BYRD, The World Bank
Structure, Trend, and Reform of Chinese Industrial Prices
CHUAN-LUN WANG, Chinese People's University
Chinese Economic Reform and Its Price Policies in Recent Years
Discussants: LOREN BRANDT, University of Toronto
CALLA WIEMER, University of Hawaii
WINSTON CHANG, State University of New York-Buffalo

8:00 A.M. IMPLICATIONS OF THE THIRD WORLD DEBT (Joint Session with the National Economic Association)
Presiding: ROBERT BROWNE, Subcommittee for International Development, U.S. House of Representatives
Papers: WILLIAM A. DARITY, JR., University of North Carolina-Chapel Hill
Is to Forgive (the Debt) Divine?
JEFFREY SACHS, Harvard University and National Bureau of Economic Research
Debt as an Obstacle to Growth
DOROTHY M. SOBEL AND EDWARD J. FRYDL, Federal Reserve Bank of New York
Economic Prospects of the Debtor Countries
Discussants: SUE COLLINS, Harvard University
RONALD JOHNSON, Northeastern University
RICHARD FEINBERG, Overseas Development Council

8:00 A.M. NEW FRONTIERS IN TRANSPORTATION AND PUBLIC UTILITIES ECONOMICS (Joint Session with the Transportation and Public Utilities Group)
Presiding: GEORGE WILSON, Indiana University
Papers: DOUGLAS CAVES AND LAURITS CHRISTENSEN, Christensen Associates, AND KATHLEEN KUESTER, Federal Reserve Board
The Relative Efficiency of New Entrant and Incumbent Airlines Under Deregulation
THEODORE E. KEELER, University of California-Berkeley, AND JOHN S. YING, University of Delaware
Statistical Analysis of Structural Change in the Motor Carrier Industry
CLIFFORD NOWELL, Weber State University, AND JOHN TSCHERHART, University of Wyoming
Testing Theories of Regulatory Behavior
Discussants: ANDREW F. DAUGHETY, University of Iowa
IAN GALE, University of Wisconsin-Madison

10:15 A.M. AMERICAN INCOME INEQUALITY
Presiding: ROBERT D. REISCHAUER, The Brookings Institution
Papers: SHELDON DANZIGER, University of Michigan, AND EUGENE SMOLENSKY, University of California-Berkeley, AND PETER GOTTSCHALK, Boston College
Recent Developments in the Upper Tail of the Income Distribution
EDWARD M. GRAMLICH, University of Michigan, RICHARD KASTEN AND FRANK SAMMARTINO, U.S. Congressional Budget Office
Fiscal Retrenchment and Redistribution
DAVID ELLWOOD, Harvard University, AND FRANK LEVY, University of Maryland
Trends in the Inequality of Permanent Incomes
Discussants: KATHARINE L. BRADBURY, Federal Reserve Bank of Boston
STANLEY LEBERGOTT, Wesleyan University

10:15 A.M. PRIVATIZATION POLICY FOR A NEW PRESIDENT
Presiding: DAVID LINOWES, University of Illinois-Urbana-Champaign
Papers: THOMAS GALE MOORE, U.S. Council of Economic Advisers
Deregulation, Privatization, and the Market
DAVID F. LINOWES, University of Illinois-Urbana-Champaign
Privatization for Our Times
JAMES C. MILLER, U.S. Office of Management and Budget
Privatization: Challenge and Opportunity
Discussants: ROBERT CRANDALL, The Brookings Institution
STUART BUTLER, Heritage Foundation

10:15 A.M. THE DEBT PROBLEM AND INEQUALITY IN DEVELOPING COUNTRIES
Presiding: E. WAYNE NAFZIGER, Kansas State University
Papers: E. WAYNE NAFZIGER, Kansas State University
External Disequilibrium, Debt, and Inequality: The Case of Africa
IRMA ADELMAN, University of California-Berkeley, AND EDWARD J. TAYLOR, University of California-Davis
Adjustment Policies and Inequality: The Case of Mexico
ERIC THORBECKE, University of California-Berkeley
The Debt Problem, Stabilization, and Structural Adjustment and Their Socioeconomic Impact
Discussant: SAYRE P. SCHATZ, Temple University

10:15 A.M. EVALUATION OF STRUCTURAL ECONOMETRIC MODELS
Presiding: LAWRENCE R. KLEIN, University of Pennsylvania
Papers: ROBERT SHILLER, Yale University
Model Comparisons and Appraisals I
IGNAZIO VISCO, Banca d'Italia
Model Comparisons and Appraisals II
Discussants: R. JEFFREY GREEN, Indiana University
BERT HICKMAN, Stanford University
RAY FAIR, Yale University
STEPHEN MCNEES, Federal Reserve Bank of Boston

10:15 A.M. GLOBAL ECONOMIC INTEGRATION AND NATIONAL ECONOMIC POLICY
Presiding: HOWARD M. WACHTEL, American University
Papers: BENJAMIN J. COHEN, Tufts University
U.S. Policy and Third World Debt
MICHAEL AHO, Council on Foreign Relations
Whither U.S. Trade Policy
JOHN WILLOUGHBY, American University
Global Economic Integration and National Economy Policy
Discussant: SIR ALEC CAIRNCROSS, St. Peter's College, Oxford

10:15 A.M. BANKING AND MONETARY POLICY
Presiding: DAVID H. SMALL, Federal Reserve Board
Papers: GARY A. DEMSKI, University of Southern California
Banking and Monetary Control Under Deregulation
GIKAS A. HARDOUVELIS, Barnard College, Columbia University, and Federal Reserve Bank of New York
Monetary Policy Games, Inflationary Bias, and Openness
DONALD D. HESTER, University of Wisconsin-Madison
The Effects on Banking Structure and Interregional Funds Flows of an Increasing Minimum Primary Capital Requirement
Discussants: MARK GERTLER, University of Wisconsin-Madison
DAVID H. SMALL, Federal Reserve Board
STUART GREENBAUM, Northwestern University

10:15 A.M. ECONOMIC EDUCATION: FACTORS AFFECTING STUDENT LEARNING
Presiding: MICHAEL K. SALEMI, University of North Carolina-Chapel Hill
Papers: JULIA A. HEATH, Memphis State University
An Econometric Model of the Role of Gender in Economic Education
SHERRIE L. W. RHINE, Winthrop College
The Effects of State Mandates on Student Performance
MICHAEL WATTS AND GERALD J. LYNCH, Purdue University
International TA's, Textbooks, Semester Length and Other Factors Affecting Student Learning in Principles of Economics Courses
Discussants: THOMAS KNIESNER, University of North Carolina-Chapel Hill
GEORGE TAUCHEN, Duke University
MICHAEL K. SALEMI, University of North Carolina-Chapel Hill

10:15 A.M. PRODUCTIVITY ANALYSIS
Presiding: ALI DOGRAMACI, Rutgers University
Papers: PAUL ROMER, University of Chicago and University of Rochester
Aggregate Productivity Trends Viewed from a Historic Perspective
DALE W. JORGENSEN, Harvard University
Productivity and U.S. Economic Growth: 1979–1985
W. ERWIN DIEWERT, University of British Columbia
Conceptual Issues in the Measurement of Productivity in the Service Sector
LESTER C. THUROW, Massachusetts Institute of Technology
Productivity in the Service Sector
Discussants: V. KERRY SMITH, North Carolina State University-Raleigh
LAWRENCE J. LAU, Stanford University
MICHAEL DENNY, University of Toronto
ROY EPSTEIN, University of Illinois-Chicago

10:15 A.M. PRINCIPLES TO GUIDE MACROECONOMIC POLICY IN THE COMING YEARS
Presiding: OLIVIER J. BLANCHARD, Massachusetts Institute of Technology
Papers: ASSAR LINDBECK, University of Stockholm, AND DENNIS SNOWER, University of London
Macroeconomic Policy and Rent-Creating Activity in the Labor Market
MANCUR OLSON, University of Maryland
An Unfatalistic Approach to Macroeconomic Policy That Is Grounded in Microeconomic Realities
LAWRENCE H. SUMMERS, Harvard University
Multiple Equilibria and the Case Against Policy Rules
Discussants: DONALD MCCLOSKEY, University of Iowa
CHARLES L. SCHULTZE, The Brookings Institution
MURRAY WEIDENBAUM, Washington University-St. Louis

10:15 A.M. PERSPECTIVES ON THE AIRLINE INDUSTRY: A DECADE AFTER DEREGULATION
Presiding: ELIZABETH E. BAILEY, Carnegie Mellon University
Papers: SEVERIN BORENSTEIN, University of Michigan, AND PETER C. REISS, Stanford University
Price Discrimination in the U.S. Airline Industry
NANCY L. ROSE, Massachusetts Institute of Technology
Financial Influences on Airline Safety Performance
DAVID CARD, Princeton University
The Effect of Deregulation on Employment and Wages of Pilots and Flight Attendants
Discussants: ALFRED E. KAHN, Cornell University
CLIFFORD WINSTON, The Brookings Institution

10:15 A.M. THE 1987 STOCK MARKET CRASH (Joint Session with American Finance Association)
Presiding: JOSEPH GRUNDFEST, U.S. Securities and Exchange Commission
Papers: SANFORD GROSSMAN, Princeton University
Program Trading and Stock Market Volatility
GEORGE PERRY, The Brookings Institution
Regulating Markets When We Know So Little
MYRON SCHOLES, Stanford University
The Crash of 1987: The Interrelation of Markets and Market Mechanisms
Discussants: JACK COFFEE, Columbia University
ROGER KUBARYCH, New York Stock Exchange
KEN LEHN, U.S. Securities and Exchange Commission

10:15 A.M. TAX POLICY FOR THE NEXT ADMINISTRATION
Presiding: JOSEPH E. STIGLITZ, Stanford University
Papers: MARTIN FELDSTEIN, Harvard University and National Bureau of Economic Research
Taxes and Tax Reform
ALICE M. RIVLIN, The Brookings Institution
The Continuing Search for a Popular Tax
MICHAEL J. BOSKIN, Stanford University
Alternative Approaches to Taxation and Tax Reform
Discussants: JOSEPH A. PECHMAN, The Brookings Institution
GEORGE BREAK, University of California-Berkeley
JOHN B. SHOVEN, Stanford University

10:15 A.M. HOUSING PROSPECTS AND POLICIES (Joint Session with the American Real Estate and Urban Economic Association)
Presiding: PATRIC H. HENDERSHOTT, Ohio State University
Papers: DONALD R. HAURIN AND PATRIC H. HENDERSHOTT, Ohio State University
How Serious Is the Decline in Home Ownership?
EDGAR OLSEN, University of Virginia
Whither Low-Income Housing Policy?
REBECCA M. BLANK AND HARVEY S. ROSEN, Princeton University
What's Happening to Housing for Low-Income Households?
Discussants: BARRY BOSWORTH, The Brookings Institution
JAMES R. FOLLAIN, University of Illinois-Urbana-Champaign
JOHN WEICHER, U.S. Office of Management and Budget

10:15 A.M. REFORMS UNDER SOCIALISM: IMPLICATIONS FOR U.S. POLICY (Joint Session with the Association for Comparative Economic Studies)
Presiding: GREGORY GROSSMAN, University of California-Berkeley

Papers: Ed A. Hewett, The Brookings Institution
An Overview of Reforms in the Soviet Union, China, and Eastern Europe
Robert Dernberger, University of Michigan
Reforms in China: Implications for U.S. Policy
Franklyn D. Holzman, Tufts University
Reforms in the USSR and Eastern Europe: Implications for U.S. Policy
Discussants: Lawrence Krause, University of California-San Diego
Allen J. Lenz, U.S. Department of Commerce
Herbert S. Levine, University of Pennsylvania

10:15 A.M. THE ROLE GENDER PLAYS IN STRUCTURING LABOR MARKETS (Joint Session with the Industrial Relations Research Association)
Presiding: Charles Craypo, University of Notre Dame
Papers: William T. Dickens, University of California-Berkeley, and Kevin Lang, Boston University
Is the Relationship Between Gender and Labor Market Segmentation Changing?
Dale Belman and Paula B. Voos, University of Wisonsin-Madison
Race and Labor Market Segmentation Among Women Workers
Teresa Ghilarducci, University of Notre Dame
Gender and Economic Restructuring: The Case of South Bend, Indiana
Discussants: Paul S. Osterman, Massachusetts Institute of Technology
Bernard Elbaum, University of California-Santa Cruz

12:30 P.M. AEA/AFA JOINT LUNCHEON
Presiding: Joseph A. Pechman, The Brookings Institution
Speaker: Alan Greenspan, Board of Governors of the Federal Reserve System

2:30 P.M. FORTIETH ANNIVERSARY OF THE MARSHALL PLAN
Presiding: Charles P. Kindleberger, Massachusetts Institute of Technology
Panel: Sir Alec Cairncross, St. Peter's College, Oxford
Hollis B. Chenery, Harvard University
Jacob J. Kaplan, Consultant
Walter J. Levy, Retired

2:30 P.M. ANTITRUST POLICY
Presiding: Roger Noll, Stanford University
Panel: W. Dale Collins, Sherman and Sterling
George Hay, Cornell University
Sanford Litwack, Dewey Ballantine
Bruce Owen, Economists, Inc.
Frederick Warren-Boulton, U.S. Department of Justice

2:30 P.M. VOCATIONAL EDUCATION AND JOB TRAINING
Presiding: David A. Wise, Harvard University
Papers: John H. Bishop, Cornell University
The Productivity Consequences of What Is Learned in High School
Robert H. Meyer, National Assessment of Vocational Education
The Tradeoff Between Occupationally Specific and General Training: Theory, Evidence and Policy Implications
W. Norton Grubb and David Stern, University of California-Berkeley
The Distribution of Net Benefits from Vocational Education
Discussants: Charles S. Benson, University of California-Berkeley
Richard B. Freeman, Harvard University

2:30 P.M. THE PRIVATE NONPROFIT SECTOR IN A MIXED ECONOMY
Presiding: Burton A. Weisbrod, University of Wisconsin-Madison
Papers: Charles Clotfelter, Duke University
The Role of the Private Nonprofit Sector: Advice for the President
Estelle James, State University of New York-Stony Brook
Interactions Among Public, Private, and Nonprofit Provision of Education
Mark Pauly, University of Pennsylvania
Nonprofit Organizations and the Theory of the Firm

Discussants: SUSAN ROSE-ACKERMAN, Yale University
THEODORE BERGSTROM, University of Michigan
JAMES ANDREONI, University of Wisconsin-Madison

2:30 P.M. ROLE OF THE HETERODOX ECONOMISTS
Presiding: PAUL STREETEN, Boston University
Papers: ROBERT HEILBRONER, New School for Social Research
The Role of the Heterodox Economist
DAVID COLANDER, Middlebury College
Making Waves
LEONARD SILK, *New York Times*
Yes, Virginia, Institutionalists Do Exist
Discussants: PAUL STREETEN, Boston University
SUSAN LEE, Forbes Magazine
DON LAVOIE, George Mason University

2:30 P.M. U.S. TRADE POLICY: SHORT-TERM PRESSURES AND LONG-TERM STRATEGY
Presiding: ANNE O. KRUEGER, Duke University
Papers: ROBERT Z. LAWRENCE, The Brookings Institution
Costs of Protection: Is There a Better Way?
WILLIAM R. CLINE, Institute for International Economics
The Macroeconomic Environment and Trade Policy
ROBERT BALDWIN, University of Wisconsin
Trade Negotiations: Strategy and Agenda
Discussants: JAGDISH BHAGWATI, Columbia University
GEZA FEKETEKUTY, U.S. Trade Representatives Office
PAUL R. KRUGMAN, Massachusetts Institute of Technology and National Bureau of Economic Research

2:30 P.M. PUBLIC CHOICE AND PRESIDENTIAL ELECTIONS
Presiding: MANCUR OLSON, University of Maryland
Papers: WILLIAM NORDHAUS, Yale University
Further Perspectives on the Political Business Cycle
DENNIS C. MUELLER, University of Maryland
Public Choice and Electoral Politics
ROBERT TOLLISON, George Mason University
The 1988 Elections: A Public Choice Commentary
Discussants: RAY FAIR, Yale University
GORDON TULLOCK, University of Arizona

2:30 P.M. PROJECT LINK AT TWENTY
Presiding: JACQUES J. POLAK, International Monetary Fund
Papers: LAWRENCE R. KLEIN, University of Pennsylvania
Two Decades of LINK Research
BERT G. HICKMAN, Stanford University
Global Modeling: A Comparative Assessment of Project LINK
CHRISTIAN E. PETERSEN, IBM Italy SPA
Dynamic Games with Protectionist Strategies: An Application of the Project LINK Model
Discussants: GUY V. STEVENS, Board of Governors of the Federal Reserve System
RALPH C. BRYANT, The Brookings Institution
JOHN B. TAYLOR, Stanford University

2:30 P.M. ECONOMIC THEORY AND ECONOMIC POLICY
Presiding: JERRY GREEN, Harvard University
Papers: NANCY STOKEY, Northwestern University
Recent Advances in Economic Dynamics
JAMES JORDAN, University of Minnesota
The Economics of Accounting Information Systems
SANFORD GROSSMAN, Princeton University
Derivative Securities and the Informational World of Markets
Discussant: BENGT HOLMSTROM, Yale University

2:30 P.M. SPACE SATELLITE COMMUNICATION: THE STATE OF EMPIRICAL KNOWLEDGE
Presiding: ROBERT CRANDALL, The Brookings Institution
Papers: HARVEY LEVIN, Hofstra University
Latecomer Cost Handicap and Claimstaking in the Orbit Spectrum Resource
MEHEROO JUSSAWALLA, East-West Center
Mitigating Benefits of Innovative Technology That Offset Latecomer Costs
MOLLY MACAULEY, Resources for the Future
Substitution Elasticities in a New Frontier: Economic Trade-Offs Between Orbit and Spectrum
Discussants: CARSON AGNEW, Hughes Aircraft, Inc.
DOUGLAS WEBBINK, Federal Communications Commission

2:30 P.M. SOCIOECONOMICS
Presiding: ALICIA MUNNELL, Federal Reserve Bank of Boston
Papers: LESTER C. THUROW, Massachusetts Institute of Technology
How to Put Capitalists Back Into Capitalism
LAWRENCE SUMMERS, Harvard University
How to Empower the Future
AMITAI ETZIONI, Harvard University
How to Shift Resources Rather than Cause a Recession
Discussants: ISABEL V. SAWHILL, Urban Institute
JOHN BRANDL, University of Minnesota and Minnesota Legislator

2:30 P.M. ALLOCATIONAL AND DISTRIBUTIONAL EFFECTS OF THE TAX REFORM ACT OF 1986
Presiding: VALERIE AMERKHAIL, Deloitte, Haskins & Sells
Papers: JANE G. GRAVELLE, Congressional Research Service, AND LAURENCE J. KOTLIKOFF, Boston University and National Bureau of Economic Research
The Excess Burden of the Corporation Income Tax: A New View
RICHARD KASTEN, FRANK SAMMARTINO, AND ERIC TODER, U.S. Congressional Budget Office
Analysis of the Distributional Effects of the Tax Reform Act of 1986
DALE W. JORGENSEN, Harvard University, AND KUN-YOUNG YUN, Yonsei University
Excess Burden of Taxation in the U.S. After Reform
Discussants: YOLANDA K. HENDERSON, Federal Reserve Bank of Boston
JOSEPH MINARIK, Urban Institute
DANIEL R. FEENBERG, National Bureau of Economic Research

2:30 P.M. HIGH EUROPEAN UNEMPLOYMENT: CAUSES AND CURES
Presiding: OLIVIER J. BLANCHARD, Massachusetts Institute of Technology
Papers: DAVID T. COE, OECD
Is Hysteresis a Manufacturing Phenomenon?
FREDERIK HUIZINGA AND FABIO SCHIANTARELLI, Boston University
Insiders-Outsiders and Membership Rules: The Effects of Employment and Wage Bargaining
LINDA S. KOLE, Federal Reserve Board
The Unnatural Behavior of European Unemployment
Discussants: CHARLES L. SCHULTZE, The Brookings Institution
DENNIS SNOWER, University of London
MICHAEL C. BURDA, INSEAD, France

2:30 P.M. DISCOUNT RATES: THEORY AND PRACTICE (Joint Session with the Association of Environmental and Resources Economists)
Presiding: CHARLES W. HOWE, University of Colorado
Papers: ROBERT W. HARTMAN, U.S. Congressional Budget Office
Discounting Issues at CBO
ROBERT C. LIND, Cornell University
The Social Rate of Discount in an Open Economy
RANDOLPH M. LYON, U.S. General Accounting Office
Discounting for Benefit-Cost and Asset-Ownership Analyses
W. KIP VISCUSI, Northwestern University, AND MICHAEL J. MOORE, Duke University
Discounting the Value of Human Life
Discussants: ARTHUR G. FRAAS, U.S. Office of Management and Budget
PAUL R. PORTNEY, Resources for the Future

2:30 P.M. FEDERAL RESERVE POLICY: THE VOLCKER ERA (Joint Session with the National Association of Business Economists)
Presiding: LYLE GRAMLEY, Mortgage Bankers Association of America
Papers: JOHN D. HAWKE, JR., Arnold & Porter
Domestic Bank Regulatory Policy in the Volcker Era
ALLEN SINAI, The Boston Company and Brandeis University
Monetary Policy in a Changing Global Environment: The Volcker Years
RUDIGER DORNBUSCH, Massachusetts Institute of Technology
Volcker and Bank Policy
Discussants: MARTIN FELDSTEIN, Harvard University and National Bureau of Economic Research
HENRY KAUFMAN, Salomon Brothers
PAUL VOLCKER, Princeton University

2.30 P.M. ODE CHAPTER ADVISERS & REGIONAL DIRECTORS SESSION (JOINT SESSION WITH OMICRON DELTA EPSILON)
Presiding: C. LOUISE NELSON, Davidson College
Papers: W. WHITNEY HICKS, University of Missouri-Columbia
Socioeconomic Determinants of Fertility in Spain, 1985
ROSS C. SINGLETON, University of Puget Sound
The Economics of Vertical Restraints
CATHERINE C. ECKEL, Virginia Polytechnic Institute and State University
The Financial Fallout from Chernobyl: Risk Perceptions and Regulatory Response
TOM VELK AND A. R. RIGGS, McGill University
The Men That Corrupted Everyburg: New York's Allen and Graham and the National Grange, 1905–1911
Discussants: GEORGE T. POTTER, Ramapo College
TAHANY NAGGAR, West Chester University
ANNETTE E. MEYER, Trenton State College
JANE CLARY, College of Charleston

4:45 P.M. THE RICHARD T. ELY LECTURE
Presiding: JOSEPH A. PECHMAN, The Brookings Institution
Speaker: HENRY J. AARON, The Brookings Institution
Politics and the Professors Revisited

Thursday, December 29, 1988

8:00 A.M. THE THRIFT INDUSTRY IN TRANSITION
Presiding: JAMES BARTH, Federal Home Loan Bank Board
Papers: GEORGE BENSTON, Emory University, AND MICHAEL F. KOEHN, Principal Analysis Group, Inc.
Capital Dissipation and Regulation of Thrifts
R. DAN BRUMBAUGH, JR., AND ANDREW CARRON, First Boston
The Future of Thrifts in the Mortgage Market
EDWARD J. KANE, Ohio State University
The High Cost of Not Resolving FSLIC's Economic Insolvency
STUART GREENBAUM, Northwestern University, Y. S. CHAN, University of Southern California, AND A. V. THAKOR, University of California-Los Angeles
Fair-Priced Deposit Insurance, Capital Requirements and Incentive Compatibility
Discussants: EDWARD L. GOLDING, Federal Home Loan Bank Board
SUSAN WOODWARD, U.S. Department of Housing & Urban Development
HERBERT BAER, Federal Reserve Bank of Chicago

8:00 A.M. EXCHANGE RATES, ECONOMIC ACTIVITY, AND ECONOMIC POLICY
Presiding: GAIL E. MAKINEN, Congressional Research Service
Papers: HALI J. EDISON, Board of Governors of the Federal Reserve System
The Effects of Fiscal Policy on Exchange Rates
YOONBAI KIM, Southern Illinois University-Carbondale
Exchange Rates and Import Prices in the U.S.: A Varying Parameter Estimation of Exchange Rate Pass-Through
BRIAN J. CODY, Federal Reserve Bank of Philadelphia
Exchange Rate Intervention in the EMS: Do EMS Member Countries Follow Optimal Intervention Strategies?

Discussants: WILLIAM A. BOMBERGER, University of Florida
JERRY TEMPALKSI, U.S. Department of the Treasury
ROSS LEVINE, Board of Governors of the Federal Reserve System

8:00 A.M. THE YOUTH LABOR MARKET
Presiding: DONALD R. WILLIAMS, Kent State University
Papers: CYNTHIA L. SIPE AND JEAN BALDWIN GROSSMAN, Public/Private Ventures
Short-Term Impacts of the Summer Training and Education Program
RICHARD B. FREEMAN, National Bureau of Economic Research and Harvard University
The Youth Labor Market Under Full Employment: The Case of Massachusetts
LINDA DATCHER-LOURY, Tufts University
Marital Status and Labor Supply of Young Men
STEPHEN HILLS, Ohio State University
Race and Sex Differences in the Effects of Early Youth Unemployment on Subsequent Wages
Discussants: ANDREW KOHEN, James Madison University
ISABEL V. SAWHILL, Urban Institute
EMILY HOFFMAN, Western Michigan University
HARRY HOLZER, Michigan State University

8:00 A.M. WHAT CAN ECONOMICS CONTRIBUTE TO SOCIAL POLICY?
Presiding: ROBERT H. HAVEMAN, University of Wisconsin-Madison
Papers: LAWRENCE H. SUMMERS, Harvard University
Economics and Labor Market-Wage Policy
MICHAEL J. BOSKIN, Stanford University
Economics and Social Insurance Policy
EDWARD M. GRAMLICH, University of Michigan
Economics and Welfare-Transfer Policy
Discussants: MARTIN S. FELDSTEIN, Harvard University
ROBERT M. SOLOW, Massachusetts Institute of Technology
JAMES TOBIN, Yale University

8:00 A.M. CURRENT ISSUES IN MONETARY POLICY
Presiding: DAVID I. FAND, Wayne State University
Papers: PHILIP L. BROCK, Duke University
Reserve Requirements and the Inflation Tax
MARK TOMA, Miami University, AND A. STEVEN HOLLAND, University of Kentucky
The Effects of a "Lender of Last Resort" on Interest Rates
ELLEN L. HUGHES-CROMWICK, Clark University
Financial Deregulation and Monetary Control
Discussants: DAVID PYLE, University of California-Berkeley
PHILLIP CAGAN, Columbia University
JOHN TATOM, Federal Reserve Bank of St. Louis

8:00 A.M. PUBLIC CHOICE AND POLITICAL ECONOMY
Presiding: TYLER COWEN, University of California-Irvine
Papers: CHRISTOPHER B. COLBURN, Old Dominion University
The Changing Composition of Income Transfer Programs: A Public Choice Analysis
LINDA COHEN, AMI GLAZER, University of California-Irvine, AND ROGER NOLL, Stanford University
Politics and the Social Discount Rate
HENRY CHAPPELL, JR., University of South Carolina
Voting Behavior, Presidential Popularity and the Macroeconomy
Discussants: TYLER COWEN, University of California-Irvine
MARK ZUPAN, University of Southern California

8:00 A.M. WOMEN'S ECONOMIC ACTIVITY AND MINORITY FAMILY INCOME
Presiding: ALICE NAKAMURA, University of Alberta
Papers: JAMES S. CUNNINGHAM AND NADJA ZALOKAR, U.S. Commission on Civil Rights
The Labor Force Participation of Black Women and the Variability of Other Family Income: Evidence from the SIPP
HARRIETT ORCUTT DULEEP, U.S. Commission on Civil Rights, AND SETH SANDERS, University of Chicago
The Role of Women in the Economic Assimilation of Asian Immigrant Families
CORDELIA REIMERS, Hunter College, City University of New York
Immigration, Assimilation, and Labor Supply of Hispanic-American Wives

LARRY ORR AND STEPHEN BELL, Abt Associates
Job Training for Welfare Recipients: The AFDC Homemaker-Home Health Aide Demonstrations
Discussants: SOLOMON POLACHEK, State University of New York-Binghamton
FINIS WELCH, University of California-Los Angeles
ALICE NAKAMURA, University of Alberta

8:00 A.M. ENERGY POLICY: THE DRIFT FROM MONOPOLY
Presiding: MORRIS A. ADELMAN, Massachusetts Institute of Technology
Papers: DAVID O. WOOD, Massachusetts Institute of Technology
Can We Model the Process?
JAMES M. GRIFFIN, Texas A&M University
Issues in Energy Policy: Oil
CHARLES E. STALON, Federal Energy Regulatory Commission and University of Illinois
New Approaches in Electricity Regulation
CATHERINE GOODE ABBOTT, Enron Corporation
Creeping Deregulation in Gas: Pipelines
PAUL C. CARPENTER, Incentives Research, Inc.
How Much Gas Field Price Regulation?
OLAV BJERKHOLT, Norwegian Central Bureau of Statistics
Gas Regulation in Northwest Europe

8:00 A.M. STOCK MARKET VOLATILITY: ECONOMIC AND POLICY ISSUES (Joint Session with the Association of Managerial Economists)
Presiding: O. MAURICE JOY, University of Kansas
Papers: DARRELL DUFFIE AND KENNETH J. SINGLETON, Stanford University
Disentangling the Effects of Noise and Aggregate Economic Disturbances on Asset Price Volatility
ALBERT P. KYLE, University of California-Berkeley
How Volatile Was the Stock Market During the Week of October 19?
ROBERT SHILLER, Yale University
Interpreting Changes in Stock Market Volability
Discussants: BRUCE N. LEHMANN, Columbia University and National Bureau of Economic Research
KENNETH R. CONE, Chicago Mercantile Exchange
KENNETH FRENCH, University of Chicago

8:00 A.M. INTERFACES BETWEEN ECONOMICS AND PSYCHOLOGY: EXPERIMENTAL CONFRONTATIONS (Joint Session with the Economic Science Association)
Presiding: JOHN H. KAGEL, University of Houston
Papers: JACK KNETSCH, University of British Columbia, RICHARD THALER, Cornell University, AND DANIEL KAHNEMAN, University of California-Berkeley
An Experimental Test of the Endowment Effect and the Coase Theorem
MATTHEW J. ANDERSON AND SHYAM SUNDER, University of Minnesota
Bayes Rule and the Representativeness Heuristic: Professional Traders in Experimental Markets
RAYMOND C. BATTALIO, Texas A&M University, JOHN H. KAGEL, University of Houston, AND DON MACDONALD, Northeast Louisiana University
Tests of Competing Non-Expected Utility Models: Animal and Human Data
Discussants: ELIZABETH HOFFMAN, University of Wyoming
JOHN HALTIWANGER, University of Maryland

8:00 A.M. THE U.S. ECONOMIC OUTLOOK (Joint Session with the National Association of Business Economists)
Presiding: JAMES F. SMITH, University of Texas-Austin
Papers: JOSEPH W. DUNCAN, Dunn&Bradstreet Corporation
The Expansion Continues: Evidence from Corporate Surveys and Business Starts and Failures Data
DAVID BERSON, Federal National Mortgage Corporation
Financial Market Developments
WILLIAM C. DUNKELBERG, Temple University
How Small Business Views 1989

8:00 A.M. ISSUES IN LDC PUBLIC FINANCE (Joint with the National Tax Association-Tax Institute of America)
Presiding: ROBERT A. BOHM, University of Tennessee
Papers: JAMES R. ALM, University of Colorado, ROY BAHL, Syracuse University, AND MATTHEW N. MURRAY, University of Tennessee
Tax Structure and the Erosion of the Tax Base in Developing Countries

MICHAEL J. WASYLENKO, Syracuse University
Fiscal Decentralization in Less Developed Countries
ROBERT A. BOHM AND WILLIAM F. FOX, University of Tennessee
Financial Feasibility versus Willingness to Pay for Water in Rural Areas
Discussants: JOHANNES LINN, The World Bank
JORGE MARTINEZ, Georgia State University
ROBIN CANTOR, Oak Ridge National Laboratory

10:15 A.M. BUDGET POLICY OPTIONS FOR THE NEW PRESIDENT
Presiding: ROBERT REISCHAUER, The Brookings Institution
Papers: PAUL VAN DE WATER, U.S. Congressional Budget Office
Non-Defense Spending Options
LAWRENCE KORB, The Brookings Institution
Defense Spending Options
JOSEPH MINARIK, Urban Institute
Revenue Options
Discussants: CHARLES R. ROLL, JR., Rand Corporation
JOHN PALMER, Syracuse University

10:15 A.M. FINANCIAL DEREGULATION: NEXT STEP
Presiding: ROBERT LITAN, The Brookings Institution
Papers: PAUL HORVITZ, University of Houston
Options for Addressing the Thrift Crisis
GILLIAN GARCIA, U.S. General Accounting Office and Senate Banking Committee
Lender of Last Resort in the Wake of the Crash
HERBERT BAER, Federal Reserve Bank of Chicago
Financial Restructuring After the Crash
Discussants: ANDREW CARRON, First Boston
MARK FLANNERY, University of British Columbia
FRANKLIN EDWARDS, Columbia University

10:15 A.M. FEDERAL URBAN POLICY
Presiding: JOHN YINGER, Syracuse University
Papers: WILLIAM A. DARITY, JR., AND GEORGE MCCARTHY, University of North Carolina-Chapel Hill
Urban Schools and Racial Inequality
RONALD FERGUSON, Harvard University
Federal Economic Development Policy in the 1990s: Adapting to New State and Local Strategies
HELEN F. LADD, Duke University
Federal Assistance to Cities in the 1990s
WILHELMINA LEIGH, U.S. Congressional Budget Office
Issues in Federal Housing Policy in the 1990s
Discussant: ANITA SUMMERS, University of Pennsylvania

10:15 A.M. TAXES AND CORPORATE BEHAVIOR
Presiding: JOHN KARL SCHOLZ, Stanford University
Papers: AVNER KALAY, University of Utah
Dividends and Taxes Revisited
R. GLENN HUBBARD, Northwestern University and National Bureau of Economic Research, AND PETER C. REISS, Stanford University and National Bureau of Economic Research
Investment and Dividend Responses to Investment Opportunities: Evidence from the U.S. Petroleum Industry
MARK WOLFSON AND MYRON SCHOLES, Stanford University
Multinational Tax Planning and Location of Investment
Discussants: JOHN SHOVEN, Stanford University
WILLIAM GALE, University of California-Los Angeles
JAMES POTERBA, Massachusetts Institute of Technology and National Bureau of Economic Research

10:15 A.M. NEW SCHOOLS IN INDUSTRIAL ORGANIZATION
Presiding: F. M. SCHERER, Swarthmore College
Papers: HAROLD DEMSETZ, University of California-Los Angeles
Industrial Organization: Feisty at Fifty
ROBERT D. WILLIG, Princeton University
Integrating Industrial Organization Thought

WILLIAM G. SHEPHERD, University of Massachusetts-Amherst
The Process of Effective Competition
Discussants: MORRIS A. ADELMAN, Massachusetts Institute of Technology
OLIVER E. WILLIAMSON, Yale University
ELEANOR A. FOX, New York University

10:15 A.M. ECONOMICS OF THE EDUCATION INDUSTRY
Presiding: BRUCE JOHNSON, Buffalo State College
Papers: RICHARD J. MURNANE, Harvard University, AND RANDALL J. OLSEN, Ohio State University
Factors Affecting the Supply of Public School Teachers: New Evidence
ESTELLE JAMES, State University of New York-Stony Brook and the Brookings Institution
Determinants of Demand for and Supply of Public versus Private Education in Modern and Developing Countries
MICHAEL S. MCPHERSON, GORDON C. WINSTON, AND MORTON O. SCHAPIRO, Williams College
Effects of Federal Student Aid on the Supply of Higher Education
Discussants: SUSAN ROSE-ACKERMAN, Columbia University
ERIC HANUSHEK, University of Rochester
HENRY M. LAVIN, IFG

10:15 A.M. ECONOMICS AND THE ENVIRONMENT: POLICY PERSPECTIVES
Presiding: ALLEN KNEESE, Resources for the Future
Papers: MILTON RUSSELL, University of Tennessee-Knoxville
The Role of Economics and Economists at EPA
WALLACE OATES, Resources for the Future and University of Maryland
Role of Economic Incentives in Environmental Policy
PAUL R. PORTNEY, Resources for the Future
The Economics of Risk Management
Discussants: THOMAS TIETENBERG, Colby College
ALBERT NICHOLS, Harvard University

10:15 A.M. FEDERAL CREDIT PROGRAMS AND THE BUDGET: ADVICE FOR A NEW ADMINISTRATION
Presiding: MARVIN PHAUP, U.S. Congressional Budget Office
Papers: MICHAEL J. BOSKIN, Stanford University, AND WILLIAM G. GALE, University of California-Los Angeles
Implications of Federal Credit Programs for Government Budgeting
ELISABETH RHYNE, University of Maryland
U.S. Credit Assistance to Small Business at Home and Abroad
MARVIN PHAUP, U.S. Congressional Budget Office
Federal Credit Reform: The Next Steps
Discussants: BARRY BOSWORTH, The Brookings Institution
HERMAN LEONARD, Harvard University

10:15 A.M. PERSPECTIVES ON ECONOMIC DEVELOPMENT
Presiding: ELIANA A. CARDOSO, Tufts University
Papers: HOLLIS CHENERY, Harvard University
Industrialization and Growth
JOSEPH STIGLITZ, Stanford University
Development and the Market
RUDIGER DORNBUSCH, Massachusetts Institute of Technology
Financial Factors in Economic Development
Discussants: ANNE O. KRUEGER, Duke University
STANLEY FISCHER, The World Bank

10:15 A.M. LABOR MARKET TRANSITIONS OF WOMEN
Presiding: OLIVIA MITCHELL, Cornell University
Papers: REBECCA BLANK, Princeton University
The Role of Part-Time Work in Women's Labor Market Choices Over Time
LESLIE SUNDT, University of Arizona
Women in Transition and Part-Time Employment
SHULAMIT KAHN, Boston University
Female Mobility and the Returns to Seniority: Should EEO Policy Be Concerned with Promotions?

Discussants: PAULA ENGLAND, University of Texas-Dallas
MICHAEL LEEDS, Temple University
JOSEPH ALTONJI, Northwestern University

10:15 A.M. TAX DEVELOPMENTS IN EAST ASIAN COUNTRIES
Presiding: HENRY J. AARON, The Brookings Institution
Papers: HIROMITSU ISHI, Hitotsubashi University
Tax Development in Japan
JOHN RIEW, Pennsylvania State University
Taxation and Development: The Taiwan Model
MUKUL G. ASHER, National University of Singapore
Challenges Facing the Singapore Economy: What Role for Tax Reform?
Discussants: RICHARD M. BIRD, University of Toronto
MALCOLM GILLIS, Duke University
VITO TANZI, International Monetary Fund

10:15 A.M. GIVING ECONOMIC ADVICE TO THE NEW ADMINISTRATION
Presiding: ROBERT H. NELSON, U.S. Department of the Interior
Panel: MARTIN FELDSTEIN, Harvard University
RUDOLF PENNER, Urban Institute
JAMES SCHLESINGER, Center for Strategic and International Studies
CHARLES L. SCHULTZE, The Brookings Institution
BERYL SPRINKEL, Council of Economic Advisers

10:15 A.M. FISCAL EFFECTS ON ECONOMIC ACTIVITY
Presiding: ALICE M. RIVLIN, The Brookings Institution
Papers: ROBERT E. HALL, Stanford University
Monetary Response to Fiscal Policy
EDWARD M. GRAMLICH, University of Michigan
Can There Be Subnational Fiscal Policy?
LAURENCE J. KOTLIKOFF, Boston University, AND ALAN AUERBACH, University of Pennsylvania
From Deficit Delusion to Meaningful Accounting for Fiscal Policy
Discussants: ROBERT M. SOLOW, Massachusetts Institute of Technology
DOUGLAS HOLTZ-EAKIN, Columbia University
B. DOUGLAS BERNHEIM, Northwestern University

10:15 A.M. PRIVATE FOUNDATIONS AND PUBLIC POLICY
Presiding: MARSHALL ROBINSON, City University of New York
Panel: THOMAS SCHELLING, Harvard University
ALBERT REES, Alfred P. Sloan Foundation
PETER DE JANOSI, Russell Sage Foundation
PETER BELL, Edna McConnell Clark Foundation

10:15 A.M. WORLD ECONOMIC OUTLOOK AND POLICIES
Presiding: C. FRED BERGSTEN, Institute for International Economics
Panel: JOHN WILLIAMSON, Institute for International Economics
LUIGI SPAVENTA, University of Rome
FRANCO MODIGLIANI, Massachusetts Institute of Technology
MASARU YOSHITOMI, Economic Research Institute, Tokyo

12:30 P.M. LUNCHEON IN HONOR OF NOBEL LAUREATE, ROBERT M. SOLOW
Presiding: JOSEPH A. PECHMAN, The Brookings Institution
Speaker: ALAN S. BLINDER, Princeton University

2:30 P.M. MACROECONOMIC INTERACTIONS AND THE CONDUCT OF NATIONAL POLICIES IN THE GLOBAL ECONOMY
Presiding: RALPH C. BRYANT, The Brookings Institution
Papers: JOHN HELLIWELL, University of British Columbia
Reducing International Imbalances
SIMON WREN-LEWIS, National Institute of Economic and Social Research, London, AND DAVID CURRIE, Queen Mary College, London
Empirical Evaluation of Proposals for Exchange-Rate Target Zones

DAVID VINES, University of Glasgow, Scotland
North-South Macroeconomic Interactions: Where Do We Stand?
Discussants: RACHEL MCCULLOCH, Brandeis University
HALI EDISON, Board of Governors of the Federal Reserve System
STANLEY FISCHER, The World Bank

2:30 P.M. PROGRESSIVE TAXATION RECONSIDERED
Presiding: RICHARD MUSGRAVE, University of California-Santa Cruz
Papers: WILLIAM VICKREY, Columbia University
Burden Incidence Under Tax and Monetary Policy
KARL CASE, Wellesley College
Trends of Burden Incidence in the Federal-State-Local System
HARVEY ROSEN, Princeton University
Tax Progression and Optimal Taxation
Discussants: DAVID COLANDER, Middlebury College
JOSEPH MINARIK, Urban Institute
ROBERT H. HAVEMAN, University of Wisconsin-Madison

2:30 P.M. APPLIED GENERAL EQUILIBRIUM ANALYSIS
Presiding: SHERMAN ROBINSON, U.S. Department of Agriculture
Papers: JAMES BERKOVEC AND DON FULLERTON, University of Virginia
Housing as Investment and Consumption
SHANTAYANAN DEVARAJAN AND DANI RODRICK, Harvard University
Trade Liberalization with Imperfect Competition and Scale Economies
GLENN W. HARRISON, THOMAS F. RUTHERFORD, AND IAN WOOTON, University of Western Ontario
The Economic Impact of the European Community
Discussants: ROGER H. GORDON, University of Michigan
JONATHAN EATON, University of Virginia
JOHN B. SHOVEN, Stanford University

2:30 P.M. CHINA'S ECONOMIC REFORMS
Presiding: LAWRENCE B. KRAUSE, University of California-San Diego
Papers: GARY H. JEFFERSON, Brandeis University
A Decade of Chinese Economic Reforms
BARRY J. NAUGHTON, University of California-San Diego
Large and Small Enterprises in China's Reform
WILLIAM BYRD, The World Bank
Rural Industrialization and Ownership
DAVID JAY GREEN, Hosei University at Tama
China's Economic Reforms: A Healing Process or Growing Tension?
Discussants: DWIGHT PERKINS, Harvard Institute for International Development
ROBERT DERNBERGER, University of Michigan
JAMES T. H. TSAO, U.S. International Trade Commission

2:30 P.M. POLICY ISSUES IN EXCISE TAXATION
Presiding: ERIC TODER, U.S. Congressional Budget Office
Papers: MICHAEL GROSSMAN, City University of New York Graduate School and National Bureau of Economic Research
Health Benefits of Increases in Alcohol and Cigarette Taxes
GARY YOHE, Wesleyan University
The Role of Taxes in Pollution Controls: Problems and Opportunities
FRANK SAMMARTINO AND RICHARD KASTEN, U.S. Congressional Budget Office
Discussants: CHARLES CLOTFELTER, Duke University
PHILIP J. COOK, Duke University
GERALDINE GERARDI, U.S. Department of the Treasury

2:30 P.M. PROSPECTS AND STRATEGIES FOR THE AMERICAN ECONOMY
Presiding: CHARLES WAITE, U.S. Bureau of the Census
Papers: RAY GOLDBERG, Harvard Business School
Prospects and Strategies for Agriculture
KENNETH MCLENNAN, Machine and Allied Products Institute
Prospects and Strategies for Manufacturing
FAYE DUCHIN, New York University
Analyzing Strategies for the U.S. Economy

Discussants: EILEEN APPELBAUM, Temple University
MICHAEL BORRUS, University of California-Berkeley
HENRY KELLY, U.S. Office of Technology Assessment

2:30 P.M. WORK AND WELFARE
Presiding: IRENE LURIE, State University of New York-Albany
Papers: DAVID FRIEDLANDER, Manpower Demonstration Research Corporation
Impacts of Eight Employment Programs for AFDC Recipients
RONALD MINCY, Urban Institute
Minimum Wage Increases and the Working Poor
ROBERT MOFFITT, Brown University, AND BARBARA WOLFE, University of Wisconsin
The Effect of Medicaid on the Labor Supply and AFDC Participation of Female Heads
Discussants: REBECCA M. BLANK, Princeton University
CORDELIA REIMERS, Hunter College, City University of New York
MICHAEL WISEMAN, University of California-Berkeley

2:30 P.M. *R&D* AND THE RENEWAL OF PRODUCTIVITY GROWTH
Presiding: ZVI GRILICHES, Harvard University
Panel: EDWARD DENISON, The Brookings Institution
DALE W. JORGENSEN, Harvard University
RICHARD R. NELSON, Columbia University
NATHAN ROSENBERG, Stanford University
ROBERT M. SOLOW, Massachusetts Institute of Technology

2:30 P.M. ECONOMIC FLUCTUATIONS AS COORDINATION FAILURES
Presiding: WALTER P. HELLER, University of California-San Diego
Papers: SATYAJIT CHATTERJEE AND RUSSELL COOPER, University of Iowa
Multiplicity and Fluctuations in an Imperfectly Competitive Economy with Entry and Exit
V. V. CHARI, Federal Reserve Bank of Minneapolis, LARRY E. JONES, Northwestern University, AND RUDOLFO MANUELLI, Stanford University
Moral Hazard in a Model of Strategic Coordinated Production
BRUCE GREENWALD, Bell Communications Research, AND JOSEPH STIGLITZ, Stanford University
Uncertainty, Rigidities, and the Effectiveness of Monetary Policy
Discussants: ROBERT E. HALL, Stanford University
PETER DIAMOND, Massachusetts Institute of Technology
OLIVIER J. BLANCHARD, Massachusetts Institute of Technology

2:30 P.M. EXCHANGE RATE POLICY
Presiding: WILLIAM H. BRANSON, Princeton University
Papers: PAUL R. KRUGMAN, Massachusetts Institute of Technology and National Bureau of Economic Research
De-industrialization, Re-industrialization, and the Real Exchange Rate
JOHN H. WILLIAMSON, Institute for International Economics
Why the World Economy Needs Exchange Rate Management and Policy Coordination
MARTIN S. FELDSTEIN, Harvard University and National Bureau of Economic Research
The Case Against Trying to Stabilize the Dollar
Discussants: PETER B. KENEN, Princeton University
RONALD I. MCKINNON, Stanford University
JACOB FRENKEL, International Monetary Fund and University of Chicago

2:30 P.M. PUBLIC POLICY ISSUES CONCERNING ELDERLY WOMEN
Presiding: NANCY GORDON, U.S. Congressional Budget Office
Papers: KAREN DAVIS AND BARBARA LYONS, Johns Hopkins University
The Poverty of Elderly Women: Implications for Public Policy in Coming Decades
TIMOTHY MCBRIDE, Urban Institute
Women's Retirement Behavior: Implications for Their Economic Well-Being
CHRISTINE BISHOP, Brandeis University
Transitions in the Living Arrangements of the Elderly: Implications for Financing Long-Term Care
Discussants: MARILYN MOON, American Association for Retired Professors
ALICE RIVLIN, The Brookings Institution

2:30 P.M. WHAT HAS HAPPENED TO U.S. SAVINGS AND WHY?
Presiding: WILLIAM D. NORDHAUS, Yale University
Panel: BARRY BOSWORTH, The Brookings Institution
ROGER BRINNER, Data Resources
PETER HOOPER, Federal Reserve Board
LAWRENCE SUMMERS, Harvard University

2:30 P.M. THE QUALITY OF LIFE
Presiding: THOMAS SCHELLING, Harvard University
Papers: ALBERT O. HIRSCHMAN, Princeton University
Holding Opinions vs. Having Preferences
ROBERT FRANK, Cornell University
Frames of Reference and the Quality of Life
ODED STARK, Harvard University
Altruism and the Quality of Life
Discussant: BARRY NALEBUFF, Princeton University

2:30 P.M. MARX AND ANALYTICAL MARXISM (Joint Session with the History of Economics Society)
Presiding: JOHN B. DAVIS, Marquette University
Papers: JOHN ROEMER, University of California-Davis
Why Marxism Does Not Need a Labor Theory of Value
ANTONIO CALLARI, STEVEN CULLENBERG, AND JACK AMARIGLIO, Franklin and Marshall College
Analytical Marxism: A Critical Appraisal
MICHAEL LEBOWITZ, Simon Fraser University
Dialectical versus Non-Dialectical Approaches to Marxian Economics
Discussants: GARY DYMSKI, University of Southern California
ROY ROTHEIM, Skidmore College

2:30 P.M. THE 1988 DISTINGUISHED LECTURE ON ECONOMICS IN GOVERNMENT (Joint Session with the Society of Government Economists)
Presiding: JOSEPH A. PECHMAN, The Brookings Institution
Papers: ANDREW F. BRIMMER, Brimmer & Company and the University of Massachusetts-Amherst
Central Banking and Systemic Risk in Capital Markets

4:45 P.M. PRESIDENTIAL ADDRESS
Presiding: JOSEPH A. PECHMAN, The Brookings Institution
Speaker: ROBERT EISNER, Northwestern University

5:45 P.M. BUSINESS MEETING

Friday, December 30, 1988

8:00 A.M. THE ECONOMICS OF CRIME PREVENTION
Presiding: MICHAEL K. BLOCK, United States Sentencing Commission
Papers: LUKE FROEB, U.S. Department of Justice
Plea Bargaining with Uncertainty: The Empirical Approach
JOEL WALDFOGEL, Stanford University
The Optimal Relationship Between Harms and Sentences
MICHAEL K. BLOCK, U.S. Sentencing Commission, AND VERNON E. GERETY, AT&T
Are Criminals Really Different: Some Experimental Evidence on the Difference Between Student and Prison Reactions to Remedy Penalties
Discussant: RAJ SAH, Yale University

8:00 P.M. WOMEN IN POVERTY
Presiding: ISABEL V. SAWHILL, Urban Institute
Papers: BARBARA GOLDMAN, DANIEL FRIEDLANDER, AND DAVID LONG, Manpower Demonstration Research Corporation
Employment Programs for Welfare Mothers in the 1980s
PATRICIA RUGGLES, Urban Institute
Welfare Dependency and Its Causes
MARGARET SIMMS, Joint Center for Political Studies
Urban Labor Markets and Poor Black Women Heading Families
Discussants: JODIE ALLEN, *Washington Post*
RALPH SMITH, U.S. Congressional Budget Office

8:00 A.M. EFFECTS OF THE CAPITAL GAINS TAX
Presiding: DONALD W. KEIFER, Congressional Research Service
Papers: ERIC TODER, U.S. Congressional Budget Office
Revenue Effects of Capital Gains Taxes: Recent Time-Series Evidence
DONALD W. KEIFER, Congressional Research Service
A Model of the Lock-In Effect of the Capital Gains Tax on Trading Corporate Stock
YOLANDA K. HENDERSON, Federal Reserve Bank of Boston
Allocation Effects of Capital Gains Taxation
Discussants: LAWRENCE B. LINDSEY, Harvard University
ANDREW LYON, University of Maryland
PATRIC H. HENDERSHOTT, Ohio State University

8:00 A.M. DUOPOLY, COLLUSION, AND PRODUCT RIVALRY
Presiding: GARTH SALONER, Massachusetts Institute of Technology
Papers: MARK BAGNOLI, University of Michigan, AND J. BRADLEY BARBEAU, University of Chicago
Product Line Choice in a Bertrand Duopoly
K. C. FUNG, Mount Holyoke College
Intra-Industry Trade and Implicit Collusion: The Differentiated Product Case
PAUL MESSINGER, Washington University
Product Rivalry in Technologically and Marketing Driven Industries
Discussants: MICHAEL WHINSTON, Harvard University
DAVID SCHARFSTEIN, Massachusetts Institute of Technology
PIERRE RÉGIBEAU, Northwestern University

8:00 A.M. PENSION PROSPECTS AND POLICIES
Presiding: ALAN L. GUSTMAN, Dartmouth College
Papers: JAMES H. STOCK AND DAVID A. WISE, Harvard University
Pensions and Retirement
ALAN L. GUSTMAN, Dartmouth College
Pension Benefits and Pension Policies
RICHARD A. IPPOLITO, Pension Benefit Guarantee Corporation
Pension Insurance
Discussants: SYLVESTER SCHIEBER, The Wyatt Corporation
GARY FIELDS, Cornell University
ZVI BODIE, Boston University

8:00 A.M. STATE AND LOCAL FINANCE: PROBLEMS AND PROSPECTS
Presiding: GEORGE F. BREAK, University of California-Berkeley
Papers: ROBERT P. INMAN, University of Pennsylvania
Fiscal Policy Through an Economic Transition: The Philadelphia Story
THOMAS F. POGUE AND JOHN L. SOLOW, University of Iowa
State Expenditure Response to Tax Windfalls: Implications for Theories of Fiscal Illusion
JOHN M. QUIGLEY, University of California-Berkeley
Intergovernmental Fiscal Relations in the Next Decade
Discussants: MICHAEL J. BOSKIN, Stanford University
GEORGE F. BREAK, University of California-Berkeley
DANIEL L. RUBINFELD, University of California-Berkeley

8:00 A.M. COMPARING MONETARY POLICY
Presiding: DEBORAH DANKER, Council of Economic Advisors
Papers: ROBERT HETZEL, Federal Reserve Bank of Richmond
Federal Reserve Targets and Instruments Prior to October 1979 and Now
RAYMOND LOMBRA, Pennsylvania State University
Federal Reserve Behavior: Theory vs. Practice
THOMAS CARGILL, University of Nevada, AND MICHAEL HUTCHINSON, University of California-Santa Cruz
Discussants: ANN-MARIE MEULENDYKE, Federal Reserve Bank of New York
KOICHI HAMADA, Yale University

8:00 A.M. SERVICES: ARE THEY REALLY GROWING?
Presiding: HARRY L. FREEMAN, American Express
Papers: ROBERT SUMMERS AND ALAN HESTON, University of Pennsylvania
Consumers' Demand for Services: Within and Across Countries

FRANKLIN ALLEN, GERALD R. FAULHABER, AND A. CRAIG MACKINLAY, University of Pennsylvania
Unbalanced Growth Redux: New Model, New Evidence
ZOLTAN KENESSEY, Board of Governors of the Federal Reserve System
A New Index of Service Production: A Companion to the Index of Industrial Production
Discussants: ROBERT J. GORDON, Northwestern University
R. GLENN HUBBARD, Columbia University
EDWARD N. WOLFF, New York University

8:00 A.M. INTERNATIONALIZATION OF EQUITY TRADING
Presiding: WILLIAM C. FREUND, Pace University
Papers: SEYMOUR SMIDT, Cornell University
Growth of Equity Trading in World Markets
WILLIAM C. FREUND, Pace University
Electronic Trading and Linkages in International Equity Markets
KEITH BOAST, Toronto Stock Exchange
The Toronto Cats: A Case Study of Economics and Politics in Electronic Trading
Discussants: ROBERT A. SCHWARTZ, New York University
DONALD E. WEEDEN, Weeden and Company
ROBERTA S. KARMEL, Brooklyn Law School, Kelley Drye & Warren, and New York Stock Exchange

8:00 A.M. ECONOMISTS AND NATIONAL HEALTH POLICY (Joint Session with the Health Economics Research Organization)
Presiding: HERBERT E. KLARMAN, New York University
Papers: UWE REINHARDT, Princeton University
Health Economists: Saviors or Elephants in a Porcelain Shop
DOROTHY P. RICE, University of California-San Francisco
Health Care for the Aged
KAREN DAVIS, Johns Hopkins University
National Health Insurance: A Proposal
Discussants: LOUIS ROSSITER, University of Virginia
VICTOR RODWIN, New York University
JOSH WIENER, The Brookings Institution

8:00 A.M. MIGRATION AND THE HISPANIC ECONOMY IN THE UNITED STATES (Joint Session with the Hispanic Professors of Economics and Business)
Presiding: JORGE SALAZAR-CARRILLO, Florida International University
Papers: BARRY CHISWICK, University of Illinois-Chicago
Business Outcomes of Recent Immigrants to the United States
CORDELIA REIMERS, Hunter College, City University of New York
Family Income Differences of Hispanic Immigrants
ANTONIO JORGE, RAUL MONCARZ, AND JORGE SALAZAR-CARRILLO, Florida International University
Cuban Immigrants as Entrepreneuers: Recent Developments
FRANCISCO RIVERA BATIZ, Rutgers University
Business Implications of Labor Migration Between Puerto Rico and the United States
Discussants: MARTA TIENDA, University of Chicago
LUIS LOCAY, State University of New York-Stony Brook
ROGER BETANCOURT, University of Maryland

8:00 A.M. WHITHER UNIONS? (Joint Session with the Industrial Relations Research Association)
Presiding: DANIEL J. B. MITCHELL, University of California-Los Angeles
Papers: LINDA N. EDWARDS, Queens College, City University of New York
The Outlook for Unions in the Public Sector
HENRY S. FARBER, Massachusetts Institute of Technology
The Outlook for Unions in the Private Sector
PAULA B. VOOS, University of Wisconsin, AND ADRIENNE EATON, Rutgers University
The Ability of Unions to Adapt to Innovative Workplace Arrangements
Discussants: LAWRENCE R. MISHEL, Economic Policy Institute
SHELDON FRIEDMAN, United Automobile Workers
RICHARD S. BELOUS, The Conference Board

10:15 A.M. REAGANOMICS IN RETROSPECT: LESSONS FOR A NEW ADMINISTRATION
Presiding: ALICE M. RIVLIN, The Brookings Institution

Panel: ROBERT M. SOLOW, Massachusetts Institute of Technology
MURRAY WEIDENBAUM, Washington University-St. Louis
MICHAEL J. BOSKIN, Stanford University
ISABEL V. SAWHILL, Urban Institute

10:15 A.M. CAPITAL ACCUMULATION DYNAMICS: GROWTH, TRADE, AND INSTABILITY
Presiding: DONALD J. HARRIS, Stanford University
Papers: DUNCAN FOLEY, Barnard College
Finance and Instability
LANCE TAYLOR, Massachusetts Institute of Technology
Cumulative Growth Processes
STEVE FAZZARI, Washington University-St. Louis
Investment Behavior and Instability
ROBERT BLECKER, American University
International Competition and Stagnation
Discussants: JOHN G. GURLEY, Stanford University
WILLIAM A. DARITY, JR., University of North Carolina-Chapel Hill
THOMAS E. WEISSKOPF, University of Michigan

10:15 A.M. ECONOMIC EFFECTS OF ACADEMIC RESEARCH
Presiding: DALE W. JORGENSEN, Harvard University
Papers: EDWIN MANSFIELD, University of Pennsylvania
The Social Rate of Return from Academic Research
RICHARD NELSON, Columbia University
The Connections Between Technology and Science, and the Role of the University
NATHAN ROSENBERG, Stanford University
Scientific Instrumentation and University Research
Discussants: EDWARD DENISON, The Brookings Institution
JOHN KENDRICK, George Washington University
ROLF PIEKARZ, National Science Foundation

10:15 A.M. CAUSES AND EFFECTS OF MERGERS
Presiding: RICHARD RUBACK, Harvard University
Papers: MARGARET BLAIR, The Brookings Institution
The Determinants of Mergers, An Empirical Assessment of a "Free Cash Flow" Theory of Mergers
ERNST BERNDT, ANN F. FRIEDLAENDER, AND CHRISTOPHER VELLTURO, Massachusetts Institute of Technology
Efficiency Costs and Benefits of Railroad Mergers
ROBERT H. MCGUCKIN AND STEPHEN ANDREWS, U.S. Bureau of the Census
Post Acquisition Performance of Acquired Lines of Business: Do Stock Market and Accounting Data Really Tell Different Stories?
Discussants: DON SIEGEL, National Bureau of Economic Research
HENRY MACFARLAND, U.S. Department of Justice
DENNIS C. MUELLER, University of Maryland

10:15 A.M. EFFECTIVENESS OF EDUCATIONAL RESOURCES
Presiding: ANITA A. SUMMERS, University of Pennsylvania
Papers: ERIC A. HANUSHEK, University of Rochester
Expenditure, Efficiency, and Equity in Public Schooling
HENRY M. LEVIN, Stanford University
Economics of Investment in the Educationally Disadvantaged
PAUL J. TAUBMAN, University of Pennsylvania
Role of Parental Income in Educational Attainment
Discussants: RICHARD J. MURNANE, Harvard University
ANITA A. SUMMERS, University of Pennsylvania

10:15 A.M. CURRENT ISSUES IN URBAN AND PUBLIC TRANSPORTATION
Presiding: JOHN MEYER, Harvard University
Papers: STEVEN A. MORRISON, Northeastern University, AND CLIFFORD WINSTON, The Brookings Institution

The Welfare Costs of Political Support for Public Transportation
KENNETH A. SMALL, University of California-Irvine
Can Academic Economics Improve Urban Transportation?
TED MILLER AND ARLEE RENO, Urban Institute
Traffic Jams: How Stuck Are We?
Discussants: IAN SAVAGE, Northwestern University
PHILIP VITON, Ohio State University

10:15 A.M. PITFALLS IN THE USE OF ECONOMETRICS IN THE ANALYSIS OF ECONOMIC POLICY
Presiding: MARC NERLOVE, University of Pennsylvania
Panel: JAMES HECKMAN, University of Chicago
EDWARD LEAMER, University of California-Los Angeles
CHRISTOPHER SIMS, University of Minnesota
LAWRENCE SUMMERS, Harvard University

10:15 A.M. THE NEW INSTITUTIONAL ECONOMICS
Presiding: DOUGLASS NORTH, Washington University-St. Louis
Papers: LEONID HURWICZ, University of Minnesota
Modeling Institutions
GARY LIBECAP, University of Arizona
The Demand for Regulation, Quality Control Among Large and Small Firms: The Meat Inspection Act of 1906
OLIVER E. WILLIAMSON, University of California-Berkeley
Operationalizing the New Institutional Economics: The Comparative Contracting Perspective
Discussants: BENGT HOLMSTROM, Yale University
DOUGLASS NORTH, Washington University-St. Louis
BARRY WEINGAST, Hoover Institution

10:15 A.M. CURRENT ISSUES IN SOCIAL SECURITY
Presiding: ALICIA MUNNELL, Federal Reserve Bank of Boston
Papers: HENRY AARON, BARRY BOSWORTH, AND GARY BURTLESS, The Brookings Institution
The Problem of Time Inconsistency in Social Security Financial Planning
B. DOUGLAS BERNHEIM, Stanford University
Social Security Expectations and Saving
MARJORIE HONIG AND CORDELIA REIMERS, Hunter College, City University of New York
Is It Worth Eliminating the Retirement Test?
Discussants: ROBERT M. BALL, National Academy of Social Insurance
LAWRENCE H. THOMPSON, U.S. General Accounting Office
JOHN HAMBOR, Social Security Administration

10:15 A.M. REGIONAL DEVELOPMENT AND INDUSTRIAL STRUCTURE
Presiding: KATHARINE L. BRADBURY, Federal Reserve Bank of Boston
Papers: KUHN LEE AND EDWARD L. RANCK, Mississippi Research and Development Center
The Dynamics of Regional Industrial Development, Diversification, and Policy Alternatives
DAVID E. WILDASIN, Indiana University
Fiscal Competition and Interindustry Trade
CHARLES A. DIAMOND AND CURTIS J. SIMON, Clemson University
Industrial Specialization and the Returns to Labor
Discussants: MICHAEL J. WASYLENKO, Syracuse University
ROBERT TANNENWALD, Federal Reserve Bank of Boston
PATRICIA FLYNN, Bentley College

10:15 A.M. REEXAMINATIONS OF TAX INCIDENCE
Presiding: JOHN B. SHOVEN, Stanford University
Papers: JAMES POTERBA, Massachusetts Institute of Technology
Who Will Bear the Tax Burden? Incidence Issues for the Next Decade
DANIEL FEENBERG, National Bureau of Economic Research
Is Progressivity an Artifact of Measurement?
KENNETH JUDD, Stanford University
Tax Incidence with Uncertain Taxes
Discussants: R. GLENN HUBBARD, Columbia University
JOSEPH A. PECHMAN, The Brookings Institution
LAURENCE J. KOTLIKOFF, Boston University

10:15 A.M. ECONOMICS OF THE FAMILY (Joint Session with the National Economic Association)
Presiding: MARGARET C. SIMMS, Joint Center for Political Studies
Papers: ROBERT I. LERMAN, Brandeis University
Economic Opportunities for Young Males and Family Formation
CYNTHIA REXROAT, Joint Center for Political Studies
Structure and the Status of Children
JAMES B. STEWART, University of Pennsylvania
A Comparison of the Economic Status of Women with Children in Extended and Nonextended Families
Discussants: RONALD MINCY, Urban Institute
BARBARA JONES, Prairie View University
RHONDA WILLIAMS, New School for Social Research

10:15 A.M. DEBT AND CAPITAL MARKETS: AN INTERNATIONAL PERSPECTIVE (Joint Session with the North American Economics and Finance Association)
Presiding: MAHMOOD A. ZAIDI, University of Minnesota
Papers: ROBERT EISNER AND PAUL J. PEIPER, Northwestern University
The World's Greatest Debtor Nation?
CHARLES FREEDMAN AND DAVID LONGWORTH, Bank of Canada
Debt and Capital Markets: The Canadian Experience
ALLEN SINAI, Boston Company and Economic Advisors, Inc.
Deficits, Debt, and Interest Rates Premia
Discussants: H. ROBERT HELLER, Board of Governors of the Federal Reserve System
JEFFREY D. SACHS, Harvard University

10:15 A.M. STRATEGIC ANALYSIS OF INTERNATIONAL ECONOMIC PROBLEMS (Joint Session with the Peace Science Society International)
Presiding: WALTER ISARD, Cornell University
Papers: JACK HIRSHLEIFER, University of California-Los Angeles
Models of Conflict and Strategic Analysis
SOLOMON POLACHEK, State University of New York-Binghamton
Discussants: STEVEN BRAMS, New York University
MARTIN MCGUIRE, University of Maryland

10:15 A.M. OPERATION OF THE INTERNATIONAL ECONOMIC SYSTEM (Joint Session with the Society for Policy Modeling)
Presiding: DOMINICK SALVATORE, Fordham University
Papers: JAGDISH N. BHAGWATI, Columbia University
Multilateral Trade Negotiations
WILLIAM H. BRANSON, Princeton University
U.S.-Japan Relations: Conflict or Cooperation
RUDIGER W. DORNBUSH, Massachusetts Institute of Technology
The United States and the NICs
LAWRENCE R. KLEIN, University of Pennsylvania
The Case for International Policy Coordination
RONALD I. MCKINNON, Stanford University
The Dollar Exchange Rate and U.S. Monetary Policy

The DATA SERVICE & INFORMATION GMBH and the STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES present

THE NEW DATABASE
ON YOUR
PERSONAL COMPUTER

DSI GMBH

EUROSTATUS

US-$ 500* p. year

The economic indicators of Europe, the United States and Japan

ON DISK

eurostat

The official economic database
of the EC COMMISSION:

The main economic indicators of the 12 EEC-Countries, the United States and Japan.
Monthly updating for subscribers. Historical database for the period from 1960 onwards.

For further details, please contact:
☎ 0 28 43/6 02 73
*) plus postage and packing

DATA SERVICE & INFORMATION GMBH
P.O. Box 1127
D-4134 RHEINBERG 1, WEST-GERMANY

ySTAT™

For IBM Personal Computers

Integrates

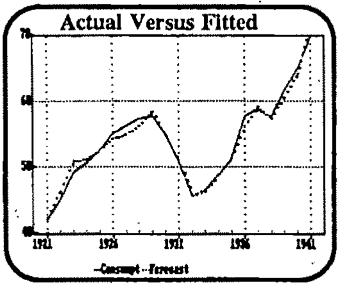
SPREADSHEET + STATISTICS + FORECASTING + GRAPHICS

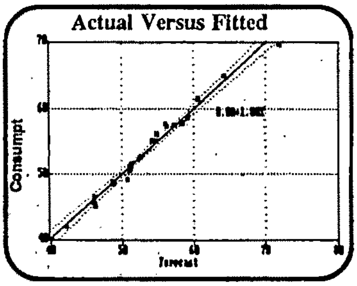
and is Easy to Use.

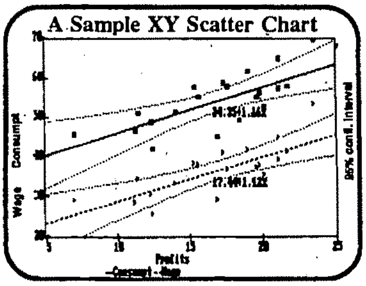
- ySTAT is the only full-fledged spreadsheet statistical program - a fast, powerful, easy-to-use stand-alone program.
- **Menu-driven** format as in Lotus 1-2-3 - offers many functions that are not available even in 1-2-3, such as vector formulas, lagged variables, dummy variables, moving averages, and interpolation of missing values.
- You can read, enter, edit, copy, move, and sort the data as well as group and partition the data for analysis right on the spreadsheet. A total of 40 function keys for one-keystroke operations.
- **Ten statistical function keys:** to tally sums and means; subtract variable means or seasonal means; standardize the data; aggregate seasonal data; generate seasonal indicators; perform moving averages, linear trend forecasts, and so on.
- **ySTAT reads 1-2-3 and Symphony worksheet files directly,** and any free-format or fixed-format textfile generated by other PC programs or downloaded from mainframe. The data can be numeric, alphabetic, including missing values.
- Mean, median, quartiles, skewness, mode, t-test, frequency distribution, correlation, crosstabs, analysis of variance.
- Multiple **regression** (OLS), standardized regression, weighted least squares, two-stage least squares, polynomial regression, and Cochrane-Orcutt method. Including diagnostic residual analysis output (see sample at right).
- **Pooling** of cross-section and time-series data.
- Two-sample difference of means test, **nonparametric tests** (Kolmogorov-Smirnov test, Wald-Wolfowitz runs test, Mann-Whitney test, Wilcoxon matched-pairs signed-ranks test) and Spearmans's and Kendall's rank-order correlations.
- **Time-series models:** autoregressive model, moving average model, and Box-Jenkins model.
- **Nonseasonal forecasts:** linear trend, quadratic trend, polynomial model, sinusoidal model; simple, double and triple exponential smoothings.
- **Seasonal forecasts:** exponential smoothings with seasonal indicators or with trigonometric functions; Winters' additive and multiplicative procedures.
- **LOGIT, PROBIT, TOBIT,** and Weighted Probit regressions: these models are twice as fast as Gauss program and faster than other programs and are much easier to use.

High-resolution **color charts** : the actual and forecast values are graphically displayed as a line chart and also as an XY chart, including a regression line and its 95% confidence interval.

```
                        OLS  -- DEPENDENT VARIABLE: Consumpt

  RIGHT-HAND          ESTIMATED          STANDARD          T_STATISTIC      PROB.
   VARIABLE          COEFFICIENT          ERROR

 1  Profits          0.192934381   (      0.09121)   T=      2.11527       0.049
 2       P-1         0.089884898   (      0.09065)   T=      0.99158       0.335
 3      W+W'         0.796218750   (      0.03994)   T=     19.93342       0.000
 4 Constant         16.236600272   (      1.30270)   T=     12.46382       0.000

  SAMPLE SIZE(  1 to   21) =               21       (DF=17)
  SUM OF SQUARED RESIDUALS =               17.879449
           VARIANCE (MSE) =                1.051732
STANDARD ERROR (ROOT MSE) =                1.025540
                R-SQUARED =                0.981008
       ADJUSTED R-SQUARED =                0.977657
    F-STATISTIC(  3,    17) =            292.707595 (p=0.0004)
        SUM OF RESIDUALS =               -0.000000
  DURBIN-WATSON STATISTIC =                1.367474

Analysis of Variance|             Source |      SUM SQ    |  DF |  MEAN SQ   |
                    |--------------------|----------------|-----|------------|
                    |Due to Regression   |        923.550 |    3|     307.850|
                    |           Residual |         17.879 |   17|       1.052|
                    |              Total |        941.430 |   20|      47.071|
                    |--------------------|----------------|-----|------------|

RESIDUAL ANALYSIS -Mean: -0.000   Adj. RMSE: 0.946   Mean Abs. % Err: 1.283

                              AUTOCORRELATIONS

LAG   COEF    T-VAL  -+-----+-----+-----+-----+-----0-----+-----+-----+-----+-----+-  LAG
  1  0.181     0.83  |                              |>>>>>>>>>>                     |   1
  2 -0.059.   -0.27  |                           <<<|                               |   2
  3 -0.033    -0.15  |                            <<|                               |   3
  4 -0.157    -0.69  |                     <<<<<<<<<|                               |   4
  5 -0.411    -1.81  |<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<|                               |   5
                     -+-----+-----+-----+-----+-----0-----+-----+-----+-----+-----+-  LAG
    Ljung-Box statistic (chi-square 4 DF): 6.698 (p=0.1527)

                          PARTIAL AUTOCORRELATION

LAG   COEF    T-VAL  -+-----+-----+-----+-----+-----0-----+-----+-----+-----+-----+-  LAG
  1  0.181     0.83  |                              |>>>>>>>>>>                     |   1
  2 -0.095    -0.44  |                        <<<<<<|                               |   2
  3 -0.004    -0.02  |                              |                               |   3
  4 -0.162    -0.74  |                    <<<<<<<<<<|                               |   4
  5 -0.378    -1.73  |<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<|                               |   5
                     -+-----+-----+-----+-----+-----0-----+-----+-----+-----+-----+-  LAG

                     ACTUAL versus FITTED VALUES and RESIDUALS

 SEQ     Actual    Fitted min=41.900                  71.873=max  Error
           (*)       (+)   -+----+----+----+----+----+----+----+-       +----+----+
   1      41.90     42.22 |#                                    |  -0.324 |   <:    |
   2      45.00     46.25 |  *+                                 |  -1.250 |<<<:    |
   3      49.20     50.77 |      *  +                           |  -1.566 |<<<:    |
   4      50.60     51.09 |        #                            |  -0.494 |   <:    |
   5      52.60     52.59 |          #                          |   0.008 |    :>   |
   6      55.10     54.23 |             +*                      |   0.869 |    :>>  |
   7      56.20     54.86 |             +  *                    |   1.338 |    :>>> |
   8      57.30     56.25 |               +*                    |   1.055 |    :>>  |
   9      57.80     58.39 |                *+                   |  -0.589 |  <<:    |
  10      55.00     54.72 |             #                       |   0.282 |    :>   |
  11      50.90     51.13 |        #                            |  -0.230 |   <:    |
  12      45.60     45.92 |  *+                                 |  -0.322 |   <:    |
  13      46.50     46.18 |   #                                 |   0.322 |    :>   |
  14      48.70     48.76 |      #                              |  -0.058 |   <:    |
  15      51.30     51.33 |         #                           |  -0.035 |   <:    |
  16      57.70     56.08 |               +*                    |   1.616 |    :>>> |
  17      58.70     59.14 |                 *+                  |  -0.436 |   <:    |
  18      57.50     57.29 |                #                    |   0.210 |    :>   |
  19      61.60     60.81 |                    +  *             |   0.989 |    :>>  |
  20      65.00     64.21 |                        +*           |   0.785 |    :>>  |
  21      69.70     71.87 |                              *    +|  -2.173 |<<<<:    |
           (*)       (+)   -+----+----+----+----+----+----+----+-       +----+----+

              LIST OF OUTLIERS - STUDENTIZED RESIDUALS GREATER THAN 1.50

   SEQ     Actual    Fitted     Residual   Std. Err  Studentized Probability
                                                      Residuals     DF=17
   ----    ------    -------    --------   --------  ----------  ----------
      3     49.20      50.77      -1.566      0.967      -1.619       0.124
     16     57.70      56.08       1.616      0.958       1.687       0.110
     21     69.70      71.87      -2.173      0.780      -2.787       0.013
   ----    ------    -------    --------   --------  ----------  ----------

 (Note: this sample output shows some of the options for regression diagnostics.)
```







MING TELECOMPUTING INC.

23 Oak Meadow Road, P.O. Box 101, Lincoln Center, MA 01773, U.S.A. (617) 259-0391

Order: ( ) ySTAT for $395. ( )ySTAT/Medical for $395.
( ) Trial Disk and information/sample output for $5.
( ) Information/sample output.
System: ( ) IBM PC. ( )XT. ( )AT. ( )System 2 Model____
( ) Other system ____________________
Options: ( ) with 8087 or 80287 or 80387 math co-processor.
( ) IBM CGA ( ) EGA ( ) Hercules Mono. Card
Type of printer or plotter ____________________

Payment: ( ) Check or money order.
( ) U.S. university or governmental purchase order.
( ) Visa. ( ) MC. Card No. ____________________
Name ____________________ Expir. date ___/___
Address ____________________
____________________ Telephone: ( ) __________

IBM is a registered trademark of International business Machines Corp.; 1-2-3 of Lotus Development Corp.; Gauss of Aptech Systems, Inc.

The National Association of Forensic Economists
Announces the

Journal of Forensic Economics

Professional journal with refereed articles on
economic loss calculations in the litigation of:

Injury/Death Anti-Trust Business Loss
Employment Discrimination

Membership includes Economists and Financial Appraisers from 50 states and over 250 Universities and Firms.

Association Membership includes subscription to the Journal of Forensic Economics, Association Newsletters and Participation in Regional and Meetings "Call for Papers" - ASSA meetings, New York, December 28 - 30, 1988

Triannual: $60.00 per year. Order from: National Association of Forensic Economists, PO Box 30067, Kansas City, MO 64112

JOB OPENINGS FOR ECONOMISTS

Available only to AEA members and institutions that agree to list their openings.

Annual Subscription Rates

U.S.A., Canada, and Mexico (first class):	$15.00, regular AEA members and institutions
	$ 7.50, junior members of AEA
All other countries (air mail):	$22.50, regular AEA members and institutions
	$15.00, junior members of AEA

Please begin my issues with:
☐ February ☐ April ☐ June ☐ August ☐ October ☐ December

Name____
First Middle Last
Address______________________________________
____
City State/Country Zip/PostalCode

Check one:
☐ I am a member of the American Economic Association.
☐ I would like to become a member. My application and payment are enclosed.
☐ (For institutions) We agree to list our vacancies in JOE.

Send payment (U.S. currency only) to:

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

Economics has more to offer than free market economics, mathematical abstrusity, and the Marxist paradigm. . .

If you DISAGREE, do not read further!

If you AGREE, consider teaching a course in your department that uses:

Economics as Social Science: An Evolutionary Approach

Wendell Gordon *The University of Texas, Austin*
and
John Adams *The University of Maryland, College Park*

Economics as Social Science is a new and modern exposition of the ideas of evolutionary (American institutional) economics. The book is a useful and stimulating introduction to the methods and theory of contemporary institutional economics. Gordon and Adams contrast the evolutionary approach with standard microeconomics and macroeconomics, and with Marxism. As a text, **Economics as Social Science** may be used in courses in non-orthodox and heterodox economics, or economic ideas.

If you are UNDECIDED, answer the following true-false questions: 1. The institutions, technology, and values of society do not change, as orthodox economic theory usually assumes. (FALSE) The world in which we live is a dynamic, changing place. 2. People are driven exclusively by rational gain considerations (profit and utility maximization), as conventional economists assume. (FALSE) Managers and consumers act on the basis of complex motives which they themselves do not fully understand. 3. Firms usually operate under conditions of perfect competition, as standard price theory alleges. (FALSE) Oligopolies are widely prevalent and it is difficult to identify even one major competitive industry in the modern American economy. 4. Economics increasingly resembles formal mathematics rather than a science of society. (TRUE) Orthodox economics uses intricate and opaque geometry, algebra, and calculus to try to prove its recondite arguments. 5. Economics is purely "positive" and "value free." (FALSE) Economics is a social science dealing with human behavior and has an obligation to explain how people's values arise and change.

About Examination Copies:

If you AGREED with the initial proposition, or if you answered correctly three or more of the true-false questions, then you are invited to request an examination copy of:

Economics as Social Science: An Evolutionary Approach
by
Wendell Gordon and John Adams

Cloth $33.00 ($25.00 for classroom use) ISBN: 0 913215 38 4

THE RIVERDALE COMPANY, Publishers
5506 Kenilworth Ave. - #102 - Riverdale, Maryland 20737 - (301) 864-2029

New Books in Business and Economics from Praeger Publishers and Quorum Books

PSYCHOLOGICAL FOUNDATIONS OF ECONOMIC BEHAVIOR
Edited by **Paul J. Albanese**
Broadens the behavioral basis of economic analysis to a social level and expands the limits of economic theory's applicability.
Available October 1988. ISBN 0-275-92742-3. $43.00 tent.

THE INVESTMENT PERFORMANCE OF CORPORATE PENSION PLANS
Why They Do Not Beat the Market Regularly
by **Stephen A. Berkowitz, Louis D. Finney,** and **Dennis E. Logue**
Examines why corporate pension plans have not matched the performance of mutual funds and what can be done to reverse this trend.
1988. ISBN 0-89930-224-6. $35.00

HANDBOOK OF THE MONEY AND CAPITAL MARKETS
by **Alan Gart**
"...thorough, comprehensive and provides an in-depth description of the financial markets and related financial instruments....a useful reference and source of guidance in investment making decisions."
—Rosanne Gatta, Assistant Vice President and Treasurer, Provident Mutual
1988. ISBN 0-89930-270-X. $49.95

POPULATION GROWTH AND SOCIOECONOMIC PROGRESS IN LESS DEVELOPED COUNTRIES
Determinants of Fertility Transition
by **Peter N. Hess**
Examines fertility rates and their trends and determinants in less developed countries by testing an empirical, interdisciplinary model of the fertility transition.
1988. ISBN 0-275-92979-5. $39.95

U.S. INTERNATIONAL COMPETITIVENESS
Evolution or Revolution?
by **John C. Hilke** and **Philip B. Nelson**
Challenges the common assumptions concerning the causes of America's growing trade deficit.
1988. ISBN 0-275-92964-7. $37.95

THE MODERN ECONOMICS OF HOUSING
A Guide to Theory and Policy for Finance and Real Estate Professionals
by **Randall Johnston Pozdena**
A highly readable examination of the various theories that have been advanced to explain the economic behavior of today's housing market.
1988. ISBN 0-89930-230-0. $39.95

SMALL BUSINESS, BANKS, AND SBA LOAN GUARANTEES
Subsidizing the Weak or Bridging a Credit Gap?
by **Elisabeth Holmes Rhyne**
Clarifies the reasons for the often heated debate about the Small Business Administration and offers new insights into whether the program does indeed subsidize the weak or perform a valuable service in bridging the small business credit gap.
1988. ISBN 0-89930-256-4. $39.95

NET ENERGY ANALYSIS AND THE ENERGY REQUIREMENTS OF ENERGY SYSTEMS
by **Daniel T. Spreng**
Thoroughly explains the theoretical principles underlying net energy analysis, offers examples of how these principles are applied, and provides an impartial critique of current methods.
1988. ISBN 0-275-92796-2. $47.95

Praeger Publishers/Quorum Books
A division and imprint of Greenwood Press, Inc.
88 Post Road West, P.O. Box 5007, Westport, CT 06881, (203) 226-3571

PRINCETON UNIVERSITY PRESS

REASONS FOR WELFARE
The Political Theory of the Welfare State
Robert E. Goodin

Robert Goodin passionately and cogently defends the welfare state from current attacks by the New Right. But he contends that the welfare state finds false friends in those on the Old Left who would justify it as a hesitant first step toward some larger, ideally just form of society.

"...erudite and very well researched, yet written in an extraordinarily clear, jargon-free, and concise style. The issues with which it deals are of central importance not only in political philosophy but also in real-world policy debates."
—Allen Buchanan, University of Arizona
Studies in Moral, Political, and Legal Philosophy
Marshall Cohen, Editor
Paper: $14.95 ISBN 0-691-02279-8
Cloth: $45.00 ISBN 0-691-07766-5

DEMOCRACY AND THE WELFARE STATE
Edited by *Amy Gutmann*

These essays explore the moral foundations and the political prospects of the welfare state in the United States. They confront powerful criticisms that have been leveled against the welfare state by conservatives, liberals, and radicals and suggest reforms in welfare state programs that might meet these criticisms.
Studies from the Project on the Federal Social Role
Forrest Chisman and Alan Pifer, Series Directors
Paper: $12.95 ISBN 0-691-02275-5
Cloth: $37.50 ISBN 0-691-07756-8

THE POLITICS OF SOCIAL POLICY IN THE UNITED STATES
Edited by *Margaret Weir, Ann Shola Orloff,* and *Theda Skocpol*

This volume places the welfare debates of the 1980s in the context of past patterns of U.S. policy, such as the Social Security Act of 1935, the failure of efforts in the 1940s to extend national social benefits and economic planning, and the backlashes against "big government" that followed reforms of the 1960s and early 1970s. The editors suggest how new family and employment policies might revitalize political coalitions and further basic national values.
Studies from the Project on the Federal Social Role
Forrest Chisman and Alan Pifer, Series Directors
Paper: $14.50 ISBN 0-691-02841-9
Cloth: $52.00 ISBN 0-691-09436-5

AT YOUR BOOKSTORE OR
Princeton University Press
41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 452-4900
ORDERS: 800-PRS-ISBN (777-4726)

This fall, a lot of micro students will exceed expectations—thanks to Eaton and Eaton's MICROECONOMICS

Eaton and Eaton is the intermediate micro text based on the notion that learning economics is doing economics. The notion is working.

Published less than one year ago, Eaton and Eaton has been adopted at a long list of schools across the country. How does a new text become so popular so rapidly? With exceptionally clear writing and careful development of ideas to help students grasp important material. Find out for yourself. Eaton and Eaton is worth a close look.

"...a well thought out book. Especially attractive is the approach emphasizing that (micro)economics involves building models and solving problems."
—Peter Huang, *Tulane University*

"A first rate text." —O. Honkalehto, *Colgate University*

1988, 471 pages, ISBN 0-7167-1818-9 With complete supplementary package

W. H. Freeman and Company
The book publishing arm of *Scientific American* • 41 Madison Avenue, New York, NY 10010

NEW FROM THE IMF

Staff Studies for the World Economic Outlook, July 1988

Two background papers prepared for the latest World Economic Outlook comprise this year's volume of staff studies. The first paper gives an analysis of the accuracy of the short-term forecasts in the Outlook; the second gives the details of MULTIMOD, the econometric model that supports the analysis of the forecasts. A basic reference for all interested in global economic analysis and forecasting.

ISBN: 1-55775-023-8

Price: US$15.00 (US$11.00 to university libraries, faculty members, and students).

Available from: Publication Services • Box No. E-413
International Monetary Fund • 700 19th Street, N.W.
Washington, D.C. 20431, U.S.A. • Telephone (202) 623-7430

Oxford

Employment Futures

Reorganization, Dislocation, and Public Policy

PAUL OSTERMAN, *Massachusetts Institute of Technology*

"Most Americans earn most of their income in the labor market. The skills in that labor market in turn are the prime determinant of any country's success or failure. Given this individual and community reality, it is important that we both understand our current labor and learn how to improve it. Paul Osterman's *Employment Futures* is an essential place to start on that process."
—Lester Thurow, *Massachusetts Institute of Technology*

1988 224 pp. $24.95

The Economics of Missing Markets, Information, and Games

Edited by FRANK HAHN, *University of Cambridge*

This collection of papers is the result of a University of Cambridge study of the consequences of missing markets, asymmetric information, market-dependent information, strategic market situations, and the role of quantity signals. The contributors also consider the behavior of overlapping generation models and their macroeconomic implications, providing a useful reference text on most of the main issues of current interest to economic theorists.

October 1988 256 pp. $55.00

The Open Economy

Tools for Policymakers in Developing Countries

Edited by RUDIGER DORNBUSCH, *Massachusetts Institute of Technology,* and
F. LESLIE C.H. HELMERS, *formerly of the World Bank*

This important book is a teaching manual, written in nontechnical language, on policymaking in developing countries. Written at the request of the Economic Development Institute of the World Bank, the papers focus on policy instruments, their use and constraints, and provide case studies of economic policy in Argentina, Brazil, Indonesia, Korea, and Mexico to illustrate basic problems and possible solutions.

(*A World Bank Publication*) (*EDI Series in Economic Development*)

1988 424 pp.; illus. cloth $38.00 paper $14.95

Keynes's Vision

ATHOL FITZGIBBONS, *Griffith University, Queensland*

In this readable yet scholarly reevaluation of the thought of John Maynard Keynes, Fitzgibbons traces the great economist's vision as it developed from his early philosophical writings (including many unpublished or neglected works) through the *Collected Writings.* Focusing on how Keynes understood significant political and economic matters, Fitzgibbons charts the evolution of his system of political economy, provides fresh insights into his approach to economic policy, and challenges the view of Keynes as a political liberal.

1988 224 pp. $44.00

Consumer Safety Regulation

Putting a Price on Life and Limb

PETER ASCH, *Rutgers University*

This book, written by an economist, critically assesses the rationales for, and the effects of, our major consumer safety programs. Incorporating relevant literature on cognitive psychology as well as economics, the author argues that although legitimate reasons for public protection of consumers exist in some markets, the particular programs we adopt often produce results that fall far short of what their advocates desire, and at least occasionally yield perverse outcomes.

1988 190 pp.; 17 linecuts $24.95

Prices and publication dates are subject to change.
To order, send check or money order to: Humanities and Social Sciences Marketing Department

Oxford University Press
200 Madison Avenue • New York, NY 10016

North-Holland
New Books in Economics

Statistical Data Bank Systems

Socio-Economic Database and Model Building in Japan

edited by **K. Uno** *and* **S. Shishido**

1988 x + 362 pages
Price: US $92.00/Dfl. 175.00
ISBN 0-444-70397-7

This book reports on the latest developments in statistical database and quantitative analysis in Japan. The articles are the outcome of the Project MUSE (Multiple-Use Socio-Economic) Data Bank. The purpose of the research was to promote interdisciplinary research regarding the feasibility of a statistical data bank in the social science field in response to increasing social needs.

Explaining the Growth of Government

edited by **J.A. Lybeck** *and* **M. Henrekson**

(Contributions to Economic Analysis, 171)

1988 viii + 396 pages
Price: US $86.75/Dfl. 165.00
ISBN 0-444-70426-4

The aim of this book is to explain the post-war growth of the public sector in a number of developed economies. The purpose is to see whether scientists familiar with their respective countries' institutional, political and economic framework, but still working as a group, can advance some common factors behind the growth of government. The authors are either economists or political scientists (or both) but having a strong empirical background.

A Study of Trade among Developing Countries, 1950-1980: An Appraisal of the Emerging Pattern

by **H.C. Thomas**

(Contributions to Economic Analysis, 170)

1988 xvi + 242 pages
Price: US $65.75/Dfl. 125.00
ISBN 0-444-70385-3

The accelerated growth of trade among developing countries during the 1970s has attracted both policy and analytical interest. While the latter interest is concerned to explain this phenomenon either in relation to existing theories of trade or through new or special theories, the former interest is concerned with the implications of this trade for global economic welfare and in particular, for the income growth and development prospects of the poorer, underdeveloped regions of the world.

Spatial Analysis and Planning under Imprecision

by **Y. Leung**

(Studies in Regional Science & Urban Economics, 17)

1988 xviii + 376 pages
Price: US $86.75/Dfl. 165.00
ISBN 0-444-70390-X

The book deals with complexity, imprecision, human valuation, and uncertainty in spatial analysis and planning, providing a systematic exposure of a new philosophical and theoretical foundation for spatial analysis and planning under imprecision. Regional concepts and regionalization, spatial preference-utility-choice structures, spatial optimization with single and multiple objectives, dynamic spatial systems and their controls are analyzed in sequence.

Plans and Disequilibria in Centrally Planned Economies

Empirical Investigations for Poland

by **W. Charemza** *and* **M. Gronicki**

(Contributions to Economic Analysis, 159)

1988 xi + 188 pages
Price: US $94.75/Dfl. 180.00
ISBN 0-444-70100-1

Co-publication with PWN Publishers. Distributed in the East European countries by Ars Polona, Warsaw, Poland

The purpose of this study is to investigate interrelations between planning mechanisms and disequilibria in a case where the planning decisions are centralized and are exogenously given to enterprises.

North-Holland

In the U.S.A. and Canada:
Elsevier Science Publishing Co. Inc.,
P.O. Box 1663, Grand Central Station,
New York, NY 10163, U.S.A.

In all other countries:
Elsevier Science Publishers,
Book Order Department,
P.O. Box 211,
1000 AE Amsterdam, The Netherlands.

US $ prices are valid only in the USA and Canada. In all other countries the Dutch Guilder (Dfl.) price is definitive. Customers in the Netherlands, please add 6% B.T.W. In New York State applicable sales tax should be added. All prices are subject to change without prior notice.

NH/ECON/BK/0970a

While the economic market may fluctuate...

the quality of our texts never does.

Just Revised

PRINCIPLES OF ECONOMICS, 4th Edition
Ryan C. Amacher and Holley H. Ulbrich
Clemson University

Featuring a balanced presentation of principles from the market perspective, this introductory package is available in comprehensive format or in split versions (one macro, one micro).

This newly revised edition covers the most important topics in micro- and macroeconomics: aggregate supply—market and government failure—international issues. The emphasis is on pragmatic analysis, not strong ideologies.

Just Revised

ECONOMICS: The Science of Common Sense, 6th Edition
Elbert V. Bowden, Chair of Banking, Appalachian State University

Providing students with a conceptual understanding of economics, this revised text features a variety of pedagogical aids that make learning and teaching simpler.

An up-to-date emphasis on conditions of the late 1980s ('87 stock market crash, government deficit, trade deficit), plus coverage of international topics in each chapter, enhances students' understanding of current economic trends and issues.

ECONOMICS: A Contemporary Introduction
William McEachern, University of Connecticut

This engaging new text covers current issues in macroeconomics (U.S. productivity, federal budgeting, and the natural rate hypothesis) as well as microeconomics (the role of time and information in production and consumption) with fresh, contemporary style.

Numerous examples, diagrams, and case studies reinforce theory throughout the text, while the supplementary package is the most complete available.

MACROECONOMICS: Theory and Policy
Steven M. Sheffrin, University of California, Davis
David A. Wilton, University of Waterloo
David M. Prescott, University of Guelph

This rigorous, forthright text juxtaposes the standard IS-LM curve with the PEP model to provide a solid framework of analysis.

Timely topics, such as hysteresis and the European unemployment rate, exchange rate overshooting, the new classical economics, and real business cycle theory, are covered in detail.

For more information on our stable economics texts, contact the South-Western representative in your area.

SOUTH-WESTERN
COLLEGE DIVISION
5101 Madison Road • Cincinnati, OH 45227

AEA sponsored Group Life Insurance for you and your family— at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA participates in a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by premium credits. In the past nine years, insured members received credits on their April 1 semiannual payment notices averaging 40% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future premium credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program
1255 23rd Street, N.W.
Washington, D.C. 20037

J-6

Please send me more information about the AEA Life Insurance Plan.

Name ______________________________ Age ________

Address ____________________________________

City ______________________ State ________ Zip ________

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

Begin making plans to attend the

Annual Meeting of The American Economic Association

(in Conjunction with Allied Social Science Associations)

to be held in

NEW YORK, NY

Dec. 27-30, 1988

The Employment Center opens Tuesday, December 27; sessions begin Wednesday, December 28.

See the September *AER* for the American Economic Association's preliminary program.

The 1989 meeting will be held in Atlanta, GA, December 27-30.

ROBERT S. / DANIEL L.
PINDYCK / RUBINFELD
MASSACHUSETTS INSTITUTE OF TECHNOLOGY / UNIVERSITY OF CALIFORNIA, BERKELEY
MICROECONOMICS

PAGE 592
CHAPTER 17

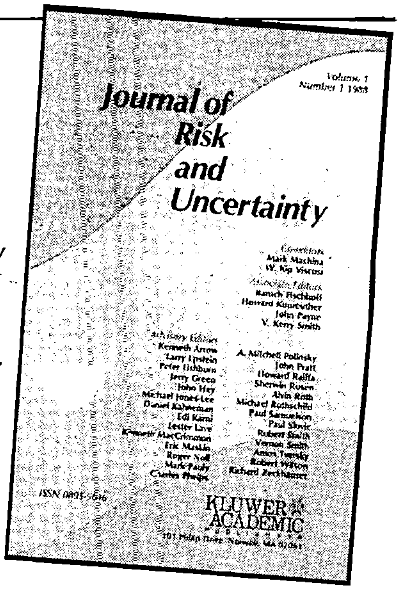
THEORY AT WORK
MACMILLAN
COLLEGE DIVISION, 866 THIRD AVE., NEW YORK, NY 10022-GIVE US A CALL AT 1-800-428-3750

An exciting new journal from Kluwer

The Journal of Risk and Uncertainty

Co-Editors: **Mark Machina**, *University of California, San Diego*, and **W. Kip Viscusi,** *Northwestern University*

Associate Editors: **Baruch Fischhoff,** *Carnegie Mellon University,* **Howard Kunreuther,** *Wharton School, University of Pennsylvania,* **John Payne,** *Duke University,* and **V. Kerry Smith,** *North Carolina State University*



Papers included in Volume 1, Number 2, June 1988

Measuring Values: A Conceptual Framework for Interpreting Transactions with Special Reference to Contingent Valuation of Visibility
Baruch Fischhoff and Lita Furby

The Marginal Value of Job Safety: A Contingent Valuation Study
Shelby Gerking, Menno de Haan and William Schulze

Consumer Processing of Hazard Warning Information
Wesley A. Magat, W. Kip Viscusi and Joel Huber

Learning About Radon's Risk
V. Kerry Smith, William H. Desvousges, Ann Fisher and F. Reed Johnson

The Journal of Risk and Uncertainty welcomes original manuscripts, both theoretical and empirical, dealing with the analysis of risk-bearing behavior and decision making under uncertainty. The topics covered in the journal include, but are not limited to:

- decision theory and the economics of uncertainty
- psychological models of choice under uncertainty
- risk and public policy
- experimental investigations of behavior under uncertainty
- empirical studies of real work risk-taking behavior

An important aim of the **JRU** is to encourage interdisciplinary communication and interaction between researchers in the area of risk and uncertainty.

Subscription Information Rates per volume; 1988 (4 issues) including postage and handling
Individual: $45.00 Institutional: $107.00 ISSN 0895-5646

Information for Authors, sample copy requests, and subscriptions should be addressed to:

KLUWER ACADEMIC PUBLISHERS
101 Philip Drive • Norwell, MA 02061

Computer Assisted Instruction

TUTOR

A COMPLETE LEARNING ENVIRONMENT

TUTOR is a tutorial software package which helps students understand concepts and receive better grades in their courses. The material of a course is divided into topics. Each topic is introduced with a statement of the lesson goal and purpose. The student may choose to obtain a printed copy of the lesson. The student may also store the lesson in a file and return to it later for analysis and evaluation of progress. The understanding of each concept is verified with a variety of questions. For each question, explanations and references to specific pages in four leading textbooks are provided. The questions are either multiple choice, true/false, fill-in, graphical or numerical applications, or short essay type. A course diskette averages 1,000 questions with explanations.

Several procedures are available:
- a learning procedure with explanations,
- a tutorial procedure with page references,
- a closed book blind test without explanation.

Additional explanations include access to
- glossary of terms,
- commonly used abbreviations,
- examples, and
- numerical applications.

The learning environment includes:
- a procedure to study and understand graphs,
- a word processor to write assignments,
- several assignments for each course,
- a built-in spreadsheet for working with data,
- relevant data sets for analysis,
- biographies of major contributors in the field,
- outlines on major research findings,
- brief summary of important laws and events.

Learning with TUTOR is dynamic.
The questions are not a boring or repetitive drill because they are virtually never exactly the same: the student cannot memorize the answer, but must learn a concept (the answers are randomized). The lessons are challenging because the student can keep track of his/her score and progress in learning. Nevertheless, the lessons are not frustrating because explanations and guidance are provided. The material is highly authoritative because it is tied to several textbooks (the student is asked to refer to these manuals). In spite of the extent of the variety of the procedures and materials available, the package is extremely user friendly and REQUIRES NO KNOWLEDGE OF COMPUTERS. Once the program is started, explanations on what to do appear at each step, and the student has access to additional explanations with a help key. The package is currently used in over one hundred colleges nationwide.

Courses and availability dates:

- Macroeconomics(immediately),
- Microeconomics (immediately),
- Basic Economics (immediately),
- Corporate Finance (Fall 1988),
- Money and Banking (Fall 1988),
- Investment (Fall 1988),
- Accounting (Spring 1989),
- Principles of Marketing (Spring 1989),
- Principles of Management (Spring 1989).

Other social science, natural science, applied science and language course packages will be ready in 1989 or 1990.

TUTOR comes in an attractive and useful package. TUTOR is priced to be affordable by all at $10.00 for each course diskette and each program diskette. The program diskette will run all the course diskettes. TUTOR is currently available for IBM and compatible with minimum of 256K memory. An Apple version will be ready in fall 1988. TUTOR is 100% money back unconditionally guaranteed.

Find TUTOR at your bookstore.Feel free to call or write us for purchasing information and technical support.

Academic Software Corporation
Bull Hill Road, Conesville, NY 12076 Tel: 1-800-666-3652

New York University Press

Agricultural Reform Efforts in the United States and Japan
D. GALE JOHNSON, Editor, University of Chicago
90 pages/ISBN: 0-8147-4171-1/$25.00

Monetary and Fiscal Policy
DOUGLAS FISHER, North Carolina State University
310 pages/ISBN: 0-8147-2587-2/$45.00

Keynes and Public Policy After Fifty Years
Volume I: Economics and Policy
Volume II: Theories and Method
OMAR F. HAMOUDA and JOHN N. SMITHIN, Editors, York University, Toronto
Volume I: 170 pages/ISBN: 0-8147-3444-8/$35.00
Volume II: 216 pages/ISBN: 0-8147-3445-6/$45.00

Marx and Keynes on Economic Recession
The Theory of Unemployment and Effective Demand
CLAUDIO SARDONI, University of Rome
152 pages/ISBN: 0-8147-7871-2/$45.00

The Political Economy of Marx, 2/e
M.C. HOWARD, University of Waterloo, Canada
J.E. KING, University of Lancaster
280 pages/ISBN: 0-8147-3452-9/$35.00, cloth
ISBN: 0-8147-3453-7/$15.00, paper

Applied Behavioral Economics, 2 Vols.
SHLOMO MAITAL, Editor, Technion-Israel Institute of Technology
The Set: 831 pages/ISBN: 0-8147-5437-6/$125.00

Nicholas Kaldor
A.P. THIRLWALL, University of Kent
360 pages/ISBN: 0-8147-8177-2/$45.00

NEW YORK UNIVERSITY PRESS
Washington Square South, New York, NY 10003

Please mention THE AMERICAN ECONOMIC REVIEW *When Writing to Advertisers*

New Titles In Industrial, International, And Comparative Economics

INDUSTRIAL POLICY OF JAPAN

edited by
Ryutaro Komiya, Masahiro Okuno, and Kotaro Suzumura
translated under the supervision of
Kazuo Sato

Japanese industrial policy has long been the focus of Western attention. While most research by foreign scholars has been conducted from a political science perspective and is primarily chronological in nature, this book uses an economic approach.

This volume represents the results of a two-year study by Japanese researchers begun in May 1982. The purpose of the study was to formulate a standard economic-theoretic analytical framework to examine pre-war conditions, post-war growth, and such recent developments as declines in productivity growth rates associated with advanced capitalist economics and the growing concern over supply side policy intervention.

Translated under the supervision of Kazuo Sato, this work combines the expertise of specialists in such fields as applied microeconomic theory, public and welfare economics, and industrial organization to present a comprehensive and current description of Japanese industrial policy.

In Paperback: $32.95
ISBN: 0-12-418651-3
Casebound: $65.00
ISBN: 0-12-418650-5
1988, 590 pages

RECENT DEVELOPMENTS IN JAPANESE ECONOMICS

edited by
Ryuzo Sato and Takashi Negishi

Covering the latest economic research in Japan, this timely volume will interest all research economists in Japanese and international economies.

August 1988, c. 245 pages
$59.95 (tentative)
ISBN: 0-12-619845-4

CHINESE ECONOMIC REFORM

How Far, How Fast?

edited by
Bruce L. Reynolds

This is a re-printing of the special September 1987 issue of the *Journal of Comparative Economics* (published by Academic Press). In October 1986, 32 scholars met for a special conference on Chinese economic reform—half of the group came from China, the rest were Americans and Europeans. Contributors to this proceedings volume are some of the most eminent experts in this field. Paper topics include *Macroeconomic Policy and Response in the Chinese Economy: The Impact of the Reform Process; Between Plan and Market: The Role of the Local Sector in Post-Mao China; Enlivening Large State Enterprises: Where is the Motive Force?;* and *China's Banking System: Current Status, Perspective on Reform.*

June 1988, c. 256 pages
$29.95 (tentative)
ISBN: 0-12-587045-0

ACADEMIC PRESS
Harcourt Brace Jovanovich, Publishers
Book Marketing Dept. #40098, 1250 Sixth Ave., San Diego, CA 92101
Call Toll Free 1-800-321-5068

©1988 by Academic Press, Inc. All Rights Reserved.

Prices subject to change without notice. GD/SJ—40098

THE CRITICS ACCLAIM
THE NEW
PALGRAVE
A DICTIONARY OF ECONOMICS

Edited by
John Eatwell, Murray Milgate, Peter Newman

"The result is an unparalleled store of information...*The New Palgrave* is a testimony to the capacity of economics to illustrate the world."

ROBERT HEILBRONER
NEW YORK REVIEW OF BOOKS

"...far surpasses any other reference work on economics."

WASHINGTON POST

"*The New Palgrave* offers today's trained economists an inexhaustible supply of instruction and provocation...here are hundreds of articles that convey the fascination and importance of the subject."

THE ECONOMIST

"...an event of great importance to all economists...a *tour de force*."

RICHARD STONE

"To attempt such an undertaking was audacious, to have carried it out is astonishing, and to have done so with so remarkable a list of authors is a tribute to the intelligence and diligence of the editors."

ALAN S. BLINDER

"...an indispensable reference tool...the topics are exhaustive."

KENNETH J. ARROW

"The list of contributors reads like a who's who in economics. The range of topics is breathtaking...it is overwhelmingly impressive."

MARTIN L. WEITZMAN

"...a remarkably ambitious and extraordinarily useful enterprise."

J.K. GALBRAITH

"It will undoubtedly become the classic for this century."

BOOKLIST/REFERENCE BOOKS BULLETIN

- 4 Volumes
- 927 Contributors
- 1,261 Subject Entries
- 4.3 Million Words
- 4,194 Pages
- 655 Biographies
- 4,000 Cross References
- Price $750

For further information, or to order, please call **800 221 2123.** In NY call collect **212 481 1334.**
Or write to the publisher, **Stockton Press,** 15 East 26th St., New York, NY 10010.
We accept VISA, MASTERCARD, DINERS, AMERICAN EXPRESS

THE ECONOMICS INSTITUTE

Gateway to Successful
Master's and Doctoral Degree Studies
in Economics, Agricultural Economics, Business and Administration

Since its establishment in 1958 under the sponsorship of the American Economic Association, the Institute has provided specialized preparatory training and orientation to over 6,000 students from over 120 countries en route to over 320 universities in the United States and other English-speaking countries.

PROGRAM FEATURES

- Six to forty-six week individualized programs.
- Intensive and comprehensive review or supplementary preparatory training in: **English, Computer Usage, Economic Theory, Mathematics, Statistics, Accounting, Management, Finance and Marketing.**
- Standardized test preparation (TOEFL, GRE, GMAT).
- Orientation to U.S. campus and community life.
- University placement assistance services.
- Certificate and Diploma Programs for short-term professional trainees.

END RESULTS for

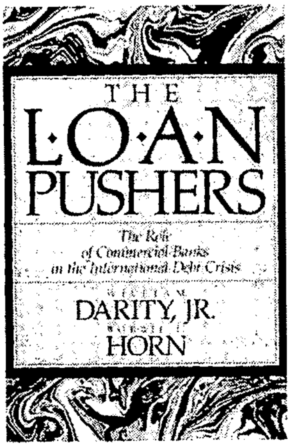
- **Associated universities:** Improved admissions procedures and a source of additional high-quality foreign students.
- **Foreign students:** Greater accessibility to U.S. universities.
- **Sponsoring organizations:** Reduced total costs.
- **And for all three:** Better grade performance records in degree programs and more rapid completion of degree requirements.

ORGANIZATION

AEA Policy and Advisory Board:
Edwin S. Mills, Northwestern University, Chairman
Lance E. Davis, California Institute of Technology
Koichi Hamada, Yale University
Joseph Havlicek, Ohio State University
Teh-wei Hu, University of California, Berkeley
Samuel A. Morley, Vanderbilt University
Ray Marshall, University of Texas at Austin
Stefan H. Robock, Columbia University

Director
Wyn F. Owen, University of Colorado

The Economics Institute, 1030 13th Street, Boulder, Colorado, 80302, USA;
Telephone: (303)492-3000; Telex: 450385 ECONINST BDR; FAX: (303)492-3006


THE
LOAN
PUSHERS
The Role of Commercial Banks in the International Debt Crisis
WILLIAM DARITY, JR.
BOBBIE L. HORN

The Loan Pushers

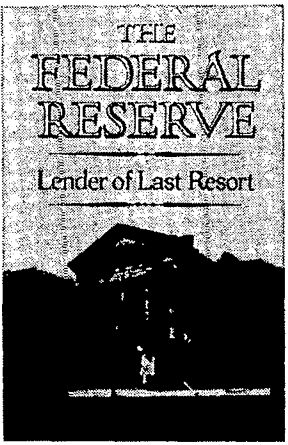
The Role of Commercial Banks in the International Debt Crisis

William Darity, Jr. and **Bobbie L. Horn**

"**The Loan Pushers** is a stunning, significant contribution to the debate swirling around events and situations having potentially grave consequences for the world's economy. The authors offer a provocative, penetrating, and at times disturbing synthesis of the controversy surrounding the genesis and implications of the global debt crisis. . . . Reading it carefully and completely is a must."

— Michael P. Claudon, Director
Middlebury Center for Economic Studies, Middlebury College

June **240 pages**
0-88730-067-7 **cloth, $29.95**



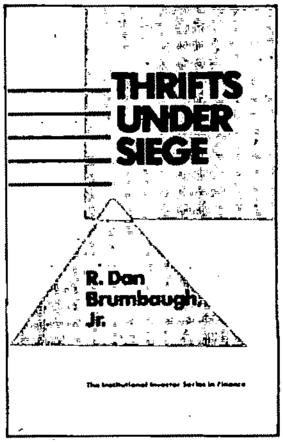
The Federal Reserve

Lender of Last Resort

Gillian Garcia and **Elizabeth Plautz**

Despite the extensive literature on bringing reform to the deficient U.S. federal financial safety net, the Federal Reserve's role as "lender of last resort" has been all but neglected. Extensively examining the first line in the government's defense against a financial crisis, **The Federal Reserve** fills the gap. The authors discuss the Fed's performance during crises since the 1970s, the impact of changes in the financial system on the lender of last resort (LLR) role, possibilities for modernizing the LLR function, and more.

November **360 pages**
0-88730-324-2 **cloth, $39.95**



Thrifts Under Siege

Restoring Order to American Banking

R. Dan Brumbaugh, Jr.

"The U.S. thrift-industry crisis may soon add to the federal budget deficit and ultimately lead to higher taxes. **Thrifts Under Siege** analyzes how the crisis developed and how it can be resolved at minimum taxpayer cost and strain to our financial system. The book is also important because it develops a long-range strategy for thrift and bank regulatory reform that is a thoughtful and significant contribution."

— David Stockman

June **240 pages**
0-88730-141-X **cloth, $39.95**

BALLINGER
PUBLISHING COMPANY

Harper & Row
Order Department
2350 Virginia Avenue
Hagerstown, MD 21740
(800) 638-3030

YES! Please send me:

____**Thrifts Under Siege** (6630024) $39.95
____**The Loan Pushers** (6610323) $29.95
____**The Federal Reserve** (6613053) $39.95

☐ Payment enclosed ☐ Bill me
Charge my ☐ MC ☐ VISA ☐ AMEX

Card no. ____________________ Exp. date________

Signature ____________________________

Send to:____________________________

____

____________________________ Zip________

My state sales tax $________
Postage and handling ($2.75/order)*$________
***Prepaid orders are postage free!**
TOTAL $________

Prices subject to change. All orders subject to credit approval. U.S. funds only. If you order by phone, tell the operator your order code is **AAER988**

FOR YOUR ECONOMICS BOOKSHELF

THE UNIVERSITY OF CHICAGO PRESS

5801 South Ellis Avenue,
Chicago, IL 60637

TRADE POLICY ISSUES AND EMPIRICAL ANALYSIS

Edited by **Robert E. Baldwin**

Ten papers address, in empirical, historical, and institutional terms, how policy shapes the effect international trade has on domestic economies.

Cloth $42.00 (est.)
352 pages (est.)
80 tables, 18 line drawings
An NBER Conference Report

WHEN PUBLIC SECTOR WORKERS UNIONIZE

Edited by **Richard B. Freeman**
and **Casey Ichniowski**

Contributors to this volume examine new data on public sector labor laws and on the labor relations practices of state and local governments.

Cloth $49.00 (est.)
416 pages (est.)
6 line drawings, 88 tables
An NBER Project Report

ECONOMIC THEORY AND THE CORE

Lester G. Telser

"Telser provides a new conceptual framework, which can lead economists to deeper insights into old problems, and to fresh insights into problems that were heretofore badly posed."—E. Roy Weintraub, *Journal of Economic Literature*

Paper $19.95 432 pages
Midway Reprint

THE RULES OF THE GAME

Reform and Evolution in the International Monetary System

Kenneth W. Dam

"Anyone interested in the history and current operations of the international monetary system will find much of value here." —Thomas D. Willett, *Journal of Economic History*

Paper $16.95 408 pages
Midway Reprint

INTERNATIONAL ECONOMIC COOPERATION

Edited by **Martin Feldstein**

International Economic Cooperation makes available the proceedings of an NBER conference organized to investigate efforts to coordinate economic policy among the developed nations.

Paper $17.95 336 pages
15 line drawings
Library cloth edition $50.00
An NBER Conference Report

ISSUES IN US-EC TRADE RELATIONS

Edited by **Robert E. Baldwin, Carl B. Hamilton,** *and* **André Sapir**

These essays by prominent North American and European trade specialists provide in-depth analyses of the economic causes of US-EC disputes and thorough consideration of the various policy options for resolving them.

Cloth $49.00
416 pages
80 tables, 18 line drawings
An NBER Conference Report

CITIES AND ECONOMIC DEVELOPMENT

From the Dawn of History to the Present

Paul Bairoch

Translated by Christopher Braider

A comprehensive study of the evolution of the city and its relation to economic life, showing how the rise of cities crucially depends at first on a community's state of agriculture and later on its state of industrialization.

Cloth $49.95
600 pages 46 tables

New

"In this one small book, protectionism — the major policy issue of our time — has been beautifully summed up by Professor Bhagwati for both expert and general reader. A 'tour de force'!" — Paul A. Samuelson

PROTECTIONISM

Jagdish Bhagwati

Clear, informative, and witty, Jagdish Bhagwati provides the best available prescription for reform in this turbulent area of trade policy. Through an appealing combination of text, quotations, cartoons, tables, charts, and graphs, Bhagwati identifies the new and powerful interests and ideologies that are likely to dominate the outcome of the debate, and offers the prognosis that the forces favoring freer trade are more robust and more fundamental than the forces of protectionism.

Protectionism is based on an inaugural series of Ohlin Lectures given at the Stockholm School of Economics in October of 1987.

Illustrated $16.95

FISCAL POLICIES AND THE WORLD ECONOMY

An Intertemporal Approach

Jacob A. Frenkel and Assaf Razin

"No serious student of international economics can afford to do without this work." — Rudiger Dornbusch

$24.00

ECONOMIC ANALYSIS OF SOLAR THERMAL ENERGY SYSTEMS

edited by Ronald E. West and Frank Kreith

This book reviews the spectrum of economic methods developed from the mid-1970s to the mid-1980s to analyze the feasibility of solar systems, and shows how the use of these techniques has influenced federally sponsored research, development, and demonstration programs.

$45.00

Now Available in Paperback

MARKETS FOR POWER

An Analysis of Electrical Utility Deregulation

Paul L. Joskow and Richard Schmalensee

"This is by far the best treatment of regulatory policy in the electric utility industry. The book succeeds in its analysis of deregulation proposals by sticking closely to economic principles, and avoiding the tiresome polemics associated with debates over competition versus government control." — Robert Weiner, *Journal of Policy Analysis and Management*

$9.95 paper ($27.50 cloth)

The MIT Press

55 Hayward Street • Cambridge, MA 02142

ACADEMIC PRESS JOURNALS

ECONOMICS

JOURNAL OF THE JAPANESE AND INTERNATIONAL ECONOMIES

Original research devoted to academic analyses of the Japanese economy and its interdependence with other national economies

Editor-in-Chief: **Masahiko Aoki**
University of Kyoto and Stanford University

Editorial Board
Michihiro Ohyama
Keio University
Koichi Hamada
Yale University
Masahiro Okuno-Fujiwara
University of Tokyo
Takatoshi Ito
University of Minnesota

Published in Cooperation with the Tokyo Center for Economic Research
Volume 2 (1988), 4 issues ISSN 0889-1583 In the U.S.A. and Canada: $100.00 All other countries: $112.00

EXPLORATIONS IN ECONOMIC HISTORY

Editor: **Larry Neal**
University of Illinois, Urbana

Volume 25 (1988), 4 issues ISSN 0014-4983
In the U.S.A. and Canada: $98.00
All other countries: $121.00

JOURNAL OF COMPARATIVE ECONOMICS

Editor: **Josef C. Brada**
Arizona State University, Tempe

The Journal of the Association for Comparative Economic Studies

Volume 12 (1988), 4 issues ISSN 0147-5967
In the U.S.A. and Canada: $110.00
All other countries: $132.00

New in 1989
GAMES AND ECONOMIC BEHAVIOR
Editor: **Ehud Kalai**
Northwestern University, Evanston, Illinois

Manuscripts may be submitted to:

Games and Economic Behavior
Department of Managerial Economics
and Decision Sciences
Leverone Hall
Northwestern University
2001 Sheridan Road
Evanston, Illinois 60208

JOURNAL OF URBAN ECONOMICS

Editor: **Edwin S. Mills**
Northwestern University, Evanston, Illinois

Volumes 23–24 (1988), 6 issues ISSN 0094-1190
In the U.S.A. and Canada: $170.00
All other countries: $206.00

Sample copies and privileged personal rates are available. For more information, please refer to no. S8162, indicate the journals that interest you, and write or call:

ACADEMIC PRESS, INC.
Journal Promotion Department
1250 Sixth Avenue, San Diego,
CA 92101, U.S.A.
(619) 230-1840

S8162

All prices are in U.S. dollars and are subject to change without notice.

PRINCETON UNIVERSITY PRESS

GALBRAITH, HARRINGTON, HEILBRONER
Economics and Dissent in an Age of Optimism
Loren J. Okroi

In a remarkably lucid style, Loren Okroi analyzes the ideas of three leading reformer-critics in the United States and places their main arguments in the context of the economic, social, and political history of postwar America. In so doing, he provides not only a skillful introduction to American social thought since the 1950s but also a wide-ranging examination of the contemporary failures of American liberal ideology.

As he explicates the works of these three men, it becomes clear that present policy debates have not even begun to resolve the dilemmas their writings have exposed.

"Okroi has made accessible to nonspecialists the major contentions of three figures whom he identifies as leading spokesmen of American democracy cum socialism.... Even for someone who has already read most of the books in question, Okroi is a useful reference."
Robert Lekachman, The City University of New York
Cloth: $29.50 ISBN 0-691-07771-1

SAVING CAPITALISM
The Reconstruction Finance Corporation and the New Deal, 1933–1940
James S. Olson

James Olson contends that the little-studied Reconstruction Finance Corporation was the major New Deal agency, even though it was the product of the Hoover Administration. Pouring more than ten billion dollars into private businesses during the 1930s in a strenuous effort to "save capitalism," the RFC was the largest, most powerful, and most influential of all New Deal agencies, proving that the main thrust of the New Deal was, in fact, state capitalism.

"Historians long have told us that the New Deal saved capitalism, but they are downright fuzzy about what it did for the capitalists because most histories omit the RFC's story. Olson's book tells this story and tells it well."
—*Jordan A. Schwartz, Northern Illinois University*
Paper: $30.00 ISBN 0-691-04749-9

AT YOUR BOOKSTORE OR
Princeton University Press
41 WILLIAM ST. • PRINCETON, NJ 08540 • (609) 452-4900
ORDERS 800-PRS-ISBN (777-4726)

AMERICAN ECONOMIC ASSOCIATION

1988 ANNUAL MEMBERSHIP RATES

Membership includes:

—a subscription to *The American Economic Review* (quarterly) plus *Papers and Proceedings*, the *Journal of Economic Literature* (quarterly) and the *Journal of Economic Perspectives* (quarterly).

● Regular members with annual incomes of $30,000 or less	$38.50
● Regular members with annual incomes above $30,000 but no more than $40,000	$46.20
● Regular members with annual incomes above $40,000	$53.90
● Junior members (available to registered students for three years only). Student status must be certified by your major professor or school registrar	$19.25
● In Countries other than the U.S.A., Add $16.00 to cover postage.	
● Family members (persons living at the same address as a regular member, additional memberships without subscription to the publications of the Association)	$7.70

Please enter my subscription for the following period:

☐ **Jan.-Dec.** ☐ **April-March** ☐ **July-June** ☐ **Oct.-Sept.**

First Name and Initial **Last Name** **Suffix**

Address Line 1

Address Line 2

City

State or Country **Zip/Postal Code**

MAJOR FIELDS (TWO ONLY)
LIST FIELDS WITH WHICH YOU CURRENTLY IDENTIFY. SELECT FIELD CODE FROM *JEL*, "Classification System for Books."

☐ ☐

Please type or print information above. Please pay with a check or money order payable in United States Dollars. Canadian and foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Please note: It is the policy of the Association, not to refund membership payments.

Endorsed by (AEA member) ____________________

Below for Junior Members Only

I certify that the person named above is enrolled as a student at ____________________

____

Authorized Signature

PLEASE SEND WITH PAYMENT TO:

AMERICAN ECONOMIC ASSOCIATION
1313 21ST AVENUE SOUTH, SUITE 809
NASHVILLE, TENNESSEE 37212-2786
U.S.A.

American Economic Association/Federal Reserve System Minority Graduate Fellowships in Economics

The American Economic Association and the Federal Reserve System are pleased to announce their joint sponsorship of graduate fellowships to minority Ph.D. students who have completed their comprehensive examinations and, if applicable, their field examinations, and are about to begin their dissertation research. Awards will be based on academic performance.

Applicants must be U.S. citizens who are Black, Hispanic or Native American and are enrolled in an accredited graduate program in Economics in the United States. Preference will be given to applicants whose area of concentration is of special interest to the Federal Reserve System (e.g., financial markets and monetary policy, nonfinancial macroeconomics, forecasting, banking markets and financial structure, regional studies, the external sector of the U.S. economy, the economies of other countries, foreign exchange markets, and international banking and financial markets).

A stipend of $700 per month for the academic year and tuition relief by the institution nominating the student and verifying successful participation in its graduate program in Economics are provided. Recipients will be assigned an adviser from the Federal Reserve System and be given the opportunity to work for one summer at the Board or a Federal Reserve Bank.

Applications are due March 1, 1989
Awards will be announced early in April 1989

For further information and application materials write:

Barbara Sears, Registrar
AEA/FRS Minority Fellowship Program
Department of Economics
College of Business and Public Administration
University of Arizona
Tucson, AZ 85721
(602) 887-5887

Japan and the World Economy

International Journal of Theory and Policy

Published in collaboration with the Center for Japan–U.S. Business and Economic Studies

Volume 1 1988/1989. First issue is scheduled to appear in October 1988.

Editor: Ryuzo Sato,
New York University, USA

Associate Editors

Hajime Hori
Tohoku University, Japan

Takashi Negishi
University of Tokyo, Japan

Rama Ramachandran
(Coordinating Editor)
Southern Methodist University and
New York University, USA

Paul Wachtel
New York University. USA

Richard Zeckhauser
Harvard University, USA

Aims and Scope

The increase in Japan's share of international trade and financial transactions has had a major impact on the world economy in general and on the U.S. economy in particular. The new economic interdependence between Japan and its trading partners created a variety of problems and so raised many issues that require further study. Japan and the World Economy will publish original research in economics, finance, managerial sciences, and marketing that express these concerns.

Japan and the World Economy seeks to meet the growing need for comprehensive, impartial and scientific study of trade imbalance and friction, technological competition, internationalization of financial markets, exchange rate variation and macroeconomic coordination, comparative managerial and marketing practices and other related topics.

Contributions are solicited from a broadbased, international community of scholars. In addition to empirical and policy-oriented studies, Japan and the World Economy also welcomes theoretical analysis that contributes to an understanding of the issues and proposes corrective action.

North-Holland
(An Imprint of Elsevier Science Publishers B.V.)

Coupon for a free specimen copy of JAPAN AND THE WORLD ECONOMY

Please send this coupon (or a photocopy) to:
Elsevier Science Publishers
Attn.: C. Fennes
P.O. Box 1991, 1000 BZ Amsterdam
The Netherlands

in the USA and Canada:
Elsevier Science Publishers
Attn.: Journal Information Center
52 Vanderbilt Avenue
New York, NY 10017, USA

Name ______________________________

Professional Address ______________________________

____

NH/ECON/JNL/A0003

The German Marshall Fund of the United States

1989 RESEARCH FELLOWSHIP PROGRAM

The German Marshall Fund of the United States is a private U.S. grant-making institution. It was established in 1972 by a gift from the Federal Republic of Germany in appreciation of American post-war recovery assistance.

Nature of Fellowships: The German Marshall Fund of the United States offers grants for research that seeks to improve the understanding of significant contemporary economic, political, and social developments involving the United States and Western Europe. Projects may focus on either comparative domestic or international issues.

Applicants: The Fund seeks to assist established U.S. scholars at various stages in their academic careers. Younger scholars, who have recently completed their studies are particularly encouraged to apply. Applicants must have completed all degree requirements by the time of application, and should have completed one or more research projects which have received critical review.

Review Process: Completed applications must be postmarked no later than **November 15, 1988.** Submissions will be reviewed by established scholars from various disciplines. An independent selection committee will make recommendations to the Fund. The Fund will announce awards by letter on **March 15, 1989.**

Approximately 12 awards will be made in 1989.

Application Forms: For application forms and additional information, contact:

The German Marshall Fund of the United States
11 Dupont Circle, N.W.
Washington, D.C. 20036
Tel: (202) 745-3950

INTERNATIONAL ECONOMIC JOURNAL

VOLUME 2 SUMMER 1988 NUMBER 2

The Effects of Fiscal Policy in Korea	Paul Evans
Effects of Direct Foreign Investment in the Presence of Sector-Specific Unemployment	Jorge G. Gonzalez
The Transfer of Programmable Automation Technology to a Rapidly Developing Economy: An Initial Assessment	Linsu Kim
Substitution among Currencies: A Preferred Habitat Hypothesis	Jagdish Handa
Interest Rates and Wage Indexation	Wookwan Rhee
International Loan to Developing Countries as a Risk Sharing Contract	Cheol-Joon Kang
The Algebra of Inflation Accounting	James Tybout
Exchange Rate Effects on Korea's Economy	Sung Y. Kwack

International Economic Journal is published quarterly in English by the Korea International Economic Association. The personal subscription price of a volume (which includes postage) is $30.00 per year. The institutional subscription price is $40.00 per year. Cheques should be made payable to the International Economic Journal and sent directly to the Managing Editor. (Volume 1 is available.)

MANUSCRIPTS SUBMITTED *(in triplicate) for publication from Asia should be sent to the Managing Editor, Professor Wontack Hong, Department of International Economics, College of Social Sciences, Seoul University, Seoul 151-742, Korea. Manuscripts from countries outside Asia should be sent to the Co-Editor, Professor Young Chin Kim, Department of Economics, Northern Illinois University, Dekalb, Illinois 60115, U.S.A.*

Of course he loves you for your mind.

But where was he when all you had on your mind was an idea demanding attention?

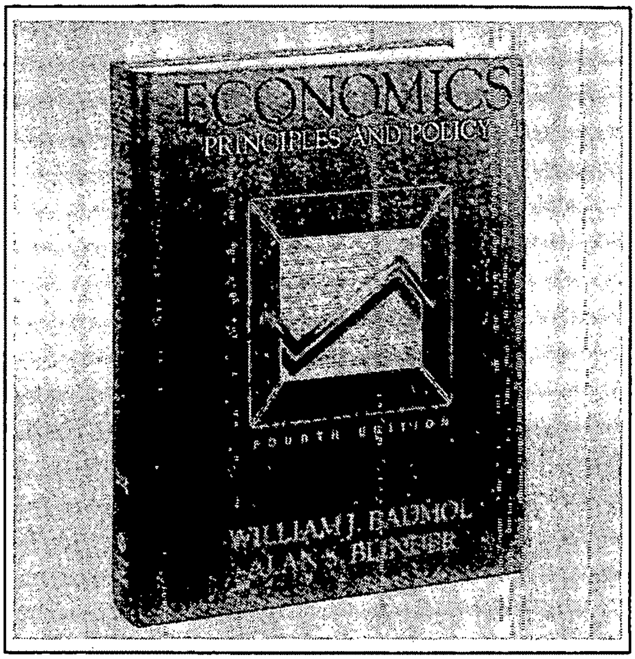
At South-Western we respect a good idea as much as a finished manuscript. And we'll work with you from conception to completion.

All along the way, you can make the most of our highly professional, in-house editors. Expert outside reviewers. And our aggressive, motivated sales force.

Call (513)527-6384. We're eager to help. Upfront, where it really counts.

SOUTH-WESTERN
COLLEGE DIVISION

Thank you
for once again making
WILLIAM J. BAUMOL AND
ALAN S. BLINDER'S



ECONOMICS
Principles and Policy
Fourth Edition
the best selling modern
principles of economics textbook.

Available in hardcover, 926 pages, and in
two-volume paperbound format:
MACROECONOMICS, 482 pages, and
MICROECONOMICS, 518 pages.
Full ancillary package available.

Coming This Fall

MICROECONOMICS
Theory and Applications
Second Edition
Fred R. Glahe and Dwight R. Lee
Hardcover, 550 pages (tent.)
Available November 1988

INTRODUCTORY ECONOMETRICS WITH APPLICATIONS
Ramu Ramanathan
Hardcover, 500 pages (tent.)
Available October 1988

COMPARING ECONOMIC SYSTEMS
A Political Economy Approach
Second Edition
Andrew Zimbalist, Howard Jay Sherman, and Stuart S. Brown
Hardcover, 480 pages (tent.)
Available October 1988

HBJ

Harcourt Brace Jovanovich, Inc.
College Sales Office
7555 Caldwell Avenue
Chicago, IL 60648
(312) 647-8822

Available Now

ECONOMICS
Private and Public Choice
Fourth Edition
James Gwartney and
Richard Stroup
Also available in two-volume format.

INTERMEDIATE MICROECONOMICS
Robert W. Clower, Phillip E. Graves, and Robert L. Sexton

THE EVOLUTION OF ECONOMIC THOUGHT
Fourth Edition
Jacob Oser and Stanley L. Brue

LABOR AND THE ECONOMY
Second Edition
Howard M. Wachtel

The Economic Forecast is
BASIC

The Great U-Turn
Corporate Restructuring and
the Polarizing of America
BENNETT HARRISON & BARRY BLUESTONE
$19.95

Rendezvous with Reality
The American Economy
after Reagan
MURRAY WEIDENBAUM
$19.95

Memoirs of an Unregulated Economist
GEORGE J. STIGLER
Sloan Foundation Science Series
$17.95

Liability
The Legal Revolution and
Its Consequences
PETER W. HUBER
$19.95

Social Science
in Government
Uses and Misuses
RICHARD P. NATHAN
$18.95

Now in Paperback!

Bailout
An Insider's Account of
Bank Failures and Rescues
IRVINE S. SPRAGUE
$9.95

Manufacturing Matters
The Myth of the
Post-Industrial Economy
STEPHEN S. COHEN & JOHN ZYSMAN
$9.95

The Transformation of
American Industrial Relations
THOMAS KOCHAN, HARRY KATZ,
& ROBERT MCKERSIE
$11.95

Unheard Voices
Labor and Economic Policy
in a Competitive World
RAY MARSHALL
$9.95

New Text Editions

Open Economy
Macroeconomics
Second Edition
RUDIGER DORNBUSCH
$24.95

Principles of Money, Banking,
and Financial Markets
Sixth Edition
LAWRENCE S. RITTER & WILLIAM L. SILBER
$36.95

At your bookstore or direct from the publisher at 1-800-638-3030.
Major credit cards accepted. Libraries and institutions call 1-800-242-7737.

BASIC BOOKS **BASIC BOOKS, INC.** 10 E. 53rd St., New York 10022

ECRS DIRECTORY OF UNPUBLISHED RESEARCH - BUSINESS & ECONOMICS
BASE VOLUME

A UNIQUE AND INVALUABLE SOURCE OF RESEARCH, DATA, INFORMATION, AND IDEAS.

Indexes over 43,000 **unpublished** works such as working / discussion / technical / seminar papers and many other types of research in **all areas of business and economics**, from the USA and abroad.

The **DIRECTORY** indexes: (a) author, (b) title, (c) affiliation, (d) I.D.#, (e) year, (f) type (g) subject category. It is organized by subject.

Dimensions - 2,000 pages; 143 subject areas; 650 institutional sources; up to 20 years; 160 work types; USA and world-wide; 9"x12"x3.

143 subject areas on a cross-section of **theory, institutions and applications** in micro/macro economics, policy, accounting, finance, marketing, O.R., engineering, law, applied math, stats, credit, econometrics, regulation, R&D, industry/country studies, monetary/fiscal topics, political economy, history, agriculture, bibliographies, computers, education, demography, insurance, women, poverty, pension, socialism, minorities, environment, energy, and many, many more (request complete list on your letter-head).

The number of institutions and individuals **registering research** on an ever increasing range of topics with ECRS is steadily rising. SUPPLEMENTS, with some 10,000 additional items, will be made available to registered BASE VOLUME purchasers at a fraction of the current price. The **BASE VOLUME will remain the only comprehensive cummulative source.**

Price: $128, plus $11 for UPS and handling. **AEA member discount: 50%.** Payment must accompany orders. Make checks payable to "ECRS, INC."

Address orders and other correspondence to:
ECONOMIC CONSULTING & RESEARCH SYSTEMS - ECRS INC
200 Park Ave - P.O. Box 1278
New York, NY 10163-1278

New from

Cambridge University Press

Horizontalists and Verticalists

The Macroeconomics of Credit Money

Basil J. Moore

Challenging the validity of much of mainstream macroeconomics, Basil Moore argues that the money supply in modern economies is not under the control of central banks, but rather is determined by borrower demand for bank credit. He then explores the implications of this perception for conventional macroeconomic theory.

About $44.50

The Underground Economies

Edgar L. Feige

The essays in this volume examine the problems of defining, measuring and understanding the implications of the "underground economies" that are surmised to have emerged in many of the world's developed nations.

About $39.50

Evolutionary Economics

Applications of Schumpeter's Ideas

Horst Hanusch, Editor

This volume contains eleven papers given at the 1986 founding meeting of the *International Joseph A. Schumpeter Society* in Augsburg. The essays raise new questions and offer additional statistical evidence to further stimulate interest and discussion about the kinds of intuitive ideas that Schumpeter introduced in his seminal period before World War I.

About $44.50

The Shapley Value

Essays in Honor of Lloyd S. Shapley

Alvin E. Roth, Editor

Composed in honor of the 65th birthday of Lloyd Shapley, this volume makes accessible the large body of work that has grown out of Shapley's seminal 1953 paper. Each of the twenty essays concerns some aspect of the Shapley value.

About $44.50

Government Policy Towards Industry in the United States and Japan

John B. Shoven, Editor

The essays contained in this volume represent state-of-the-art thinking on many important issues relating to government policies in the Japanese and American economies.

About $42.50

The Popperian Legacy in Economics

Papers Presented at a Symposium in Amsterdam, December 1985

Neil de Marchi, Editor

This volume is the first to examine just why, and in what context, Sir Karl Popper's view that science can only comprise efforts to refute empirical hypotheses has found appeal among economists.

$37.50

Equilibrium Business Cycle Theory in Historical Perspective

Kyun Kim

This book presents a historical investigation of the theoretical development of contemporary Equilibrium Business Cycle Theory (EBCT). The author examines the central features of the EBCT by tracing both the history of business cycle theory and the history of econometrics.

Historical Perspectives on Modern Economics

About $34.50

The Mobility of Labor and Capital

A Study in International Investment and Labor Flow

Saskia Sassen

"...probably the best book in the field that I have knowledge of in many years. Professor Sassen is one of the most innovative researchers in the area of urban political economy."

—*Manuel Castells*, University of California, Berkeley

"...an intelligent combination of theorizing and relevant data which is the mark of a good book."

—*Immanuel Wallerstein*, State University of New York, Binghamton

$34.50

Economics in Urban Conservation

Nathaniel Lichfield

The role of economics in urban conservation is still relatively underdeveloped. This innovative and important study examines the approach, method, and techniques for the systematic application of economics in the conservation of urban areas.

About $49.50

Conceptual Anomalies in Economics and Statistics

Lessons from the Social Experiment

Leland Gerson Neuberg

This book explores anomalies in the conceptual basis of social control experiments and in the foundation of statistics and economics. The author concludes that social control experiments are a frequently overrated tool for social policy improvements.

About $49.50

At bookstores or order from

Cambridge University Press

32 East 57th Street, NY, NY 10022

Cambridge toll-free numbers for orders only:

800-872-7423, outside NY State, 800-227-0247,

NY State only. MasterCard and Visa accepted.

Agricultural Development Principles

Economic Theory and Empirical Evidence

Robert D. Stevens and Cathy L. Jabara

What are the food and agricultural development problems facing third world nations? Does current economic theory help accelerate development? Does it foster useful development policies? The authors address these and other questions to provide a wide-ranging and thorough introduction to the theories, policies, and practices aimed at increasing food and agricultural production.

$45.00 *hardcover* **$17.50** *paperback*

Food Subsidies in Developing Countries

Costs, Benefits, and Policy Options

edited by Per Pinstrup-Andersen

Nineteen leading economists, nutritionists, and policy analysts present key findings from the International Food Policy Research Institute, synthesizing practical experiences with subsidy programs in countries throughout Asia, Africa, and Latin America. The authors emphasize lessons applicable to decisionmaking on current and future food policy for developing countries.

Published for the International Food Policy Research Institute
$39.50 *hardcover*

Land Policies and Farm Productivity in Thailand

Gershon Feder, Tongroj Onchan, Yongyuth Chalamwong,
and Chira Hongladarom

How does security of land ownership affect the behavior and performance of farmers in developing countries? Based on data gathered in Thailand, this book presents an econometric analysis of investment patterns and of differences in productivity between farmers with and without secure ownership. The results provide a frame of reference applicable to other developing countries as well.

Published for the World Bank
$18.50 *hardcover*

Analytical Tools for Sector Work in Education

Alain Mingat and Jee-Peng Tan

In this book, the first of its kind, Alain Mingat and Jee-Peng Tan provide analytical methods and arguments to help evaluate current educational policies in developing countries, identify weaknesses in the education system, and assess alternative strategies for overcoming them. They have written a practical guide for analytical work that meets an urgent need among educational policy makers, planners, and researchers.

Published for the World Bank
$25.00 *hardcover*

THE JOHNS HOPKINS UNIVERSITY PRESS
701 West 40th Street, Suite 275, Baltimore, Maryland 21211

PC-GIVE 5.0

LEAP TO THE ECONOMETRIC FRONTIER
with this FAST, FRIENDLY, POWERFUL AND INTERACTIVE
MODULAR SYSTEM
for Analysing Economic Time-Series

Ideal for TEACHING and STUDENT USE

- MENU DRIVEN WITH A BUILT-IN HELP SYSTEM
- DETAILED SELF-TEACHING TUTORIALS
- CLEAR AND COMPREHENSIVE MANUALS
- SUPERB GRAPHICS *(Crossplots; Time Series; Densities + Histograms)*
- EASY MODEL FORMULATION
- EXCELLENT SCREEN PRESENTATIONS
- RESULTS REVIEWS
- HIGH LEVEL OF ERROR PROTECTION
- SIMPLE FILE HANDLING + DOS ACCESS
- DATABANK FACILITIES

EASY TO LEARN AND EASY TO USE

yet

Ideal for ADVANCED RESEARCH

- LEAST SQUARES AND INSTRUMENTAL VARIABLES
- RECURSIVE ESTIMATION AND TESTING
- AUTOMATIC GRAPHING OF OUTPUT *(Recursive estimates + standard errors and Tests; Forecasts)*
- INTERACTIVE LINEAR SYSTEM MODELLING (2SLS, 3SLS, LIVE, FIVE, FIML+RECURSIVE VAR)
- PREPROGRAMMED DIAGNOSTIC TESTS: *(Autocorrelation; Constancy; Normality; Reset; Heteroscedasticity; Omitted Variables; ARCH)*
- DERIVED STATISTICS+IDENTIFICATION CHECKS
- COMFAC AND ENCOMPASSING TESTS
- FAST AND ACCURATE ALGORITHMS

POWERFUL AND THOROUGH ECONOMETRIC MODELLING

DEVELOPED FOR LIVE CLASSROOM DEMONSTRATIONS AND FRAUGHT LATE NIGHT RESEARCH!

PC-GIVE 5.0 NEEDS:
IBM* PC/AT or Compatible
512K RAM: 2 Disk Drives and Graphics Card (CGA, HGC)
[FIML needs 640K]
Math Co-processor recommended

PC-GIVE 5.0 COSTS:
SINGLE COPIES $380
(50% Discount Available for Accredited Academics)
SITE LICENCES ONLY
(Unlimited copies) $900

PC-GIVE is already standard at over 250 sites worldwide.
FOR FURTHER INFORMATION AND A PC-GIVE OVERVIEW, CONTACT:
PROFESSOR DAVID F. HENDRY—
OXFORD INSTITUTE OF ECONOMICS AND STATISTICS
OXFORD OX1 3UL, ENGLAND
or call GILLIAN COATES on 0144 865-271088

*IBM is a Trade Mark of International Business Machines Corporation.

Journal of International Economic Integration

Solicits Papers to Compete for the Annual Daeyang Prize in Economics ($7,000) and Welcomes Subscriptions by Interested Parties

Current Issues include

Bela Balassa, *Japanese Trade Policies Towards Developing Countries.*

Basant K. Kapur, *Open-Economy Response to a Terms of Trade Shock in a Growth Context.*

Norman C. Miller, *A Gerneral Approach to the Balance of Payments and Exchange Rates.*

Leonard F.S. Wang, *Product Market Imperfections and Customs Unions theory.*

Gene M. Grossman, *The Employment and Wage Effects of Import Competition in the United States.*

Wilfred J. Ethier and Ronald D. Fischer, *The New Protectionism*

Joshua Aizenman, *Inflation, Tariffs and Tax Enforcement Costs*

Chung H. Lee and Seiji Naya, *The Internationalization of U.S. Service Industries and Its Implications for Developing Countries.*

The Journal of International Economic Integration is published biannually (Spring and Autumn) by the Institute for International Economics, King Sejong University, Seoul, Korea.
The purpose of the Journal is to support and encourage research in the area of international trade, international finance and other related economic issues that include general professional interest in international economic affairs. Welcoming both theoretical and empirical analyses in international economics, the Journal is strongly interested in the issues of the international economic cooperation.

- The Journal welcomes unsolicited manuscripts, which will be considered for publication by the Editorial Board.
- From papers selected for publication, the Prize committee will choose the best manuscript(s) to receive the $7,000 Daeyang Prize. The winner of the prize is announced in the Spring issue every year.
- The manuscripts should be accompanied by an abstract of no more than 100 words and a brief curriculum vitae containing the author's academic career. All submissions should be typewritten, double-spaced, in English with footnotes, references, figures, tables and any other illustrative material on separate sheets.
- Three copies of the manuscript and all accompanying material should be submitted to the following address by October 31, 1988 for consideration for 1989 publication.
- For subscriptions to the Journal ($20 per year for individuals, $30 per year for institutions), send a check or money order payable to King Sejong University to the following address.

Institute for International Economics
King Sejong University
Seongdong-Ku, Seoul, Korea

The Frank E. Seidman
Distinguished Award in Political Economy
and
Rhodes College
are pleased to announce
the Recipient of the 1988 Award

Robert Triffin

Institut de Recherches Economiques, Service
de Conjuncture, Universite Catholique de Louvain, Belgium

The Award recognizes an economist who has made a unique contribution to the interdisciplinary advancement of economic thought as it applies to the implementation of public policy. Professor Triffin will receive the Award's $15,000.00 honorarium at the annual banquet in Memphis, Tennessee on September 15, 1988.

Previous recipients:

1974	Gunnar Karl Myrdal	1981	Richard A. Musgrave
1975	John Kenneth Galbraith	1982	Janos Kornai
1976	Kenneth Ewart Boulding	1983	Robert M. Solow
1977	Thomas Cromble Schelling	1984	James M. Buchanan
1978	Arthur Frank Burns	1985	Gary Stanley Becker
1979	Arthur M. Okun	1986	Amartya Kumar Sen
1980	Albert O. Hirschman	1987	William J. Baumol

Members of the 1988 Selection Committee:

James Buchanan
Irving Kristol
G. Randolph Rice
Robert M. Solow
James Tobin

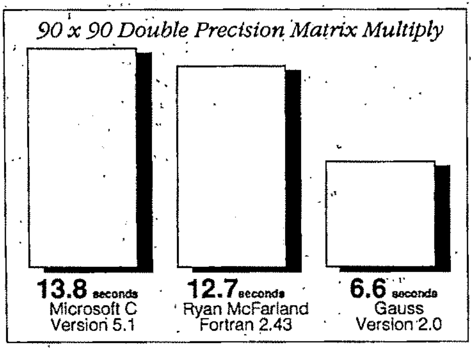
The Award is administered by a nine-member Board of Trustees comprised of eminent economists, members of the donor family, and officers and faculty of Rhodes College, Memphis, Tennessee. Director of the Award is Mel G. Grinspan.

GAUSS RUN WITH THE BEST

Saves time. Saves money. New GAUSS 2.0.

Now you can run most routines at or near the capacity of your hardware.

From Bell Labs to CalTech, research scientists and engineers have found GAUSS can help them solve problems with their PCs that would take them hours longer wth conventional languages. And not only does GAUSS run fast, it's easier to program because its language is closer to mathematical syntax. GAUSS allows you to easily enter your own ideas even if you are not a programmer and saves time by including over 250 mathematical routines. You can even call your favorite Fortran, C or assembler subroutines.



Now GAUSS 2.0, the second generation of the new breed of matrix handling programming languages is available, with indexed, easier-to-use documentation, on-line debugger and Smart™ on-line help that can even handle your own functions. And GAUSS 2.0 is faster, up to 20% on some calculations, compared to our previous versions.

The GAUSS system includes publication quality graphics and a library of mathematical routines and applications, documented in the updated 2.0 manual. All this for the total price of just $395. So don't waste another minute running at half speed, order GAUSS 2.0 today.

We're so sure GAUSS will make you more productive, giving you more answers to more questions more quickly that we guarantee it.

If you are not completely satisfied with the power, speed and flexibility of GAUSS simply return it for a full refund. No questions asked. Call **(206) 631-6679** to order.

GAUSS - faster, easier, smarter...guaranteed.

SATISFACTION • GUARANTEE 100% CUSTOMER

GAUSS-2.0

FOR IBM-PC-XT-AT-PS/2 & COMPATIBLES
WITH MATH COPROCESSOR

BY APTECH SYSTEMS, INC.
26250 - 196th Place SE
Kent Washington 98042
Phone (206) 631-6679
FAX (206) 630-1220
CALL OR WRITE FOR FREE BROCHURE

IBM IS A REGISTERED TRADEMARK OF INTERNATIONAL BUSINESS MACHINES
ALL TESTS RUN ON AN IBM PS/2 MODEL 80 16 MHz

The American Economic Review

ARTICLES

KENNETH HENDRICKS AND ROBERT H. PORTER
An Empirical Study of an Auction with Asymmetric Information

STUART E. THIEL Some Evidence on the Winner's Curse

ROBERT GIBBONS
Learning in Equilibrium Models of Arbitration

SEVERIN BORENSTEIN AND MARTIN B. ZIMMERMAN
Market Incentives for Safe Commercial Airline Operation

B. PETER PASHIGIAN
Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing

ROBERT E. MARTIN
Franchising and Risk Management

ROBERT S. PINDYCK
Irreversible Investment, Capacity Choice, and the Value of the Firm

TRACY R. LEWIS AND DAVID E. M. SAPPINGTON
Regulating a Monopolist with Unknown Demand

LAURENCE BALL AND STEPHEN G. CECCHETTI
Imperfect Information and Staggered Price Setting

CHARLES L. BALLARD
The Marginal Efficiency Cost of Redistribution

B. DOUGLAS BERNHEIM AND ODED STARK
Altruism within the Family Reconsidered: Do Nice Guys Finish Last?

RICHARD ARNOTT, ARTHUR J. HOSIOS, AND JOSEPH STIGLITZ
Implicit Contracts, Labor Mobility, and Unemployment

SHORTER PAPERS: M. D. Shapiro; L. Jonung and D. Laidler; H. I. Grossman and J. B. Van Huyck; B. W. Bateman; D. M. Cutler; T. Downs and H. Tehranian; S. Layson; C. d'Aspremont and A. Jacquemin; J. B. De Long; W. J. Baumol and E. N. Wolff; J. J. Merrick, Jr.; W. G. Dewald, J. G. Thursby, and R. G. Anderson; C. Pitchik and A. Schotter.

DECEMBER 1988

THE AMERICAN ECONOMIC ASSOCIATION

• Printed at Banta Company, Menasha, Wisconsin.

• Copyright © American Economic Association 1988. All rights reserved.

• No responsibility for the views expressed by authors in this *Review* is assumed by the editors or the publishers, The American Economic Association.

Correspondence relating to advertising, business matters, permissions to quote, subscriptions, and changes of address, should be sent to the American Economic Association, 1313 21st Avenue South, Suite 809, Nashville, TN 37212–2786. Please remit membership payment with the application included elsewhere in this journal. Change of address notice must be received at least six (6) weeks prior to the publication month. A membership or subscription paid twice is automatically extended for an additional year unless otherwise requested.

THE AMERICAN ECONOMIC REVIEW (ISSN 0002–8282), December 1988, Vol. 78, No. 5, is published five times a year (March, May, June, September, December) by the American Economic Association, 1313 21st Avenue South, Suite 809, Nashville, TN 37212–2786. Annual subscription fees: Institutional subscriber-$125.00, Individual subscriber-$72.00, Regular members-$42.00, $50.40, or $58.80 depending on income. A subscription also includes the *Journal of Economic Literature* and the *Journal of Economic Perspectives*. In countries other than the U.S.A., add $16.00 for extra postage. Second-class postage paid at Nashville, TN and at additional mailing offices. POSTMASTER: Send address changes to the *American Economic Review*, 1313 21st Avenue South, Suite 809, Nashville, TN 37212–2786.

Founded in 1885

Officers

President
ROBERT EISNER
Northwestern University

President-elect
JOSEPH A. PECHMAN
The Brookings Institution

Vice-Presidents
MARTIN S. FELDSTEIN
National Bureau of Economic Research and Harvard University
F. M. SCHERER
Swarthmore College

Secretary-Treasurer
C. ELTON HINSHAW
Vanderbilt University

Editor of The American Economic Review
ORLEY C. ASHENFELTER
Princeton University

Editor of The Journal of Economic Literature
JOHN PENCAVEL
Stanford University

Editor of The Journal of Economic Perspectives
JOSEPH E. STIGLITZ
Stanford University

Executive Committee

Elected Members of the Executive Committee
SHERWIN ROSEN
University of Chicago
THOMAS J. SARGENT
University of Minnesota
ROBERT J. BARRO
Harvard University
JUDITH A. THORNTON
University of Washington
GEORGE A. AKERLOF
University of California-Berkeley
ISABEL V. SAWHILL
Urban Institute

EX OFFICIO Members
ALICE M. RIVLIN
The Brookings Institution
GARY S. BECKER
University of Chicago

THE AMERICAN ECONOMIC REVIEW

Editor

ORLEY ASHENFELTER

Co-Editors
ROBERT H. HAVEMAN
BENNETT T. McCALLUM
HAL R. VARIAN

Production Editor
CLAIRE H. COMISKEY

Board of Editors

GEORGE A. AKERLOF
JAMES E. ANDERSON
CLAUDIA D. GOLDIN
JO ANNA GRAY
GEORGE E. JOHNSON
KENNETH L. JUDD
JOHN F. KENNAN
MAURICE OBSTFELD
EDGAR O. OLSEN
ROBERT H. PORTER
JOHN G. RILEY
RICHARD ROLL
ALVIN E. ROTH
DAVID E. M. SAPPINGTON
KENNETH J. SINGLETON
ROBERT S. SMITH
BARBARA J. SPENCER
RICHARD TRESCH
LESLIE YOUNG

- Submit manuscripts (4 copies), 50 pages maximum, single-sided, double-spaced, to: Orley Ashenfelter, Editor, *AER*; 209 Nassau Street, Princeton, NJ 08542-4607.

- Submission fee: $50 for members; $100 for nonmembers. Please pay with a check or money order payable in United States Dollars. Canadian and Foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Style guides will be provided upon request.

- Copyright © American Economic Association 1988. All rights reserved.

December 1988

VOLUME 78, NUMBER 5

Articles

An Empirical Study of an Auction with Asymmetric Information *Kenneth Hendricks and Robert H. Porter* 865

Some Evidence on the Winner's Curse *Stuart E. Thiel* 884

Learning in Equilibrium Models of Arbitration *Robert Gibbons* 896

Market Incentives for Safe Commercial Airline Operation *Severin Borenstein and Martin B. Zimmerman* 913

Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing *B. Peter Pashigian* 936

Franchising and Risk Management *Robert E. Martin* 954

Irreversible Investment, Capacity Choice, and the Value of the Firm *Robert S. Pindyck* 969

Regulating a Monopolist with Unknown Demand *Tracy R. Lewis and David E. M. Sappington* 986

Imperfect Information and Staggered Price Setting *Laurence Ball and Stephen G. Cecchetti* 999

The Marginal Efficiency Cost of Redistribution *Charles L. Ballard* 1019

Altruism within the Family Reconsidered: Do Nice Guys Finish Last? *B. Douglas Bernheim and Oded Stark* 1034

Implicit Contracts, Labor Mobility, and Unemployment *Richard Arnott, Arthur J. Hosios, and Joseph Stiglitz* 1046

Shorter Papers

The Stabilization of the United States Economy: Evidence from the Stock Market *Matthew D. Shapiro* 1067

Are Perceptions of Inflation Rational? Some Evidence from Sweden *Lars Jonung and David Laidler* 1080

Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation *Herschel I. Grossman and John B. Van Huyck* 1088

G. E. Moore and J. M. Keynes: A Missing Chapter in the History of the Expected Utility Model *Bradley W. Bateman* 1098

Tax Reform and the Stock Market: An Asset Price Approach *David M. Cutler* 1107

Predicting Stock Price Responses to Tax Policy Changes *Thomas Downs and Hassan Tehranian* 1118

Third-Degree Price Discrimination, Welfare and Profits: A Geometrical Analysis *Stephen Layson* 1131

Cooperative and Noncooperative *R&D* in Duopoly with Spillovers *Claude d'Aspremont and Alexis Jacquemin* 1133

Productivity Growth, Convergence, and Welfare:
- Comment *J. Bradford De Long* 1138
- Reply *William J. Baumol and Edward N. Wolff* 1155

Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*:
- Comment *John J. Merrick, Jr.* 1160
- Reply *William G. Dewald, Jerry G. Thursby, and Richard G. Anderson* 1162

Erratum

Honesty in a Model of Strategic Information Transmission: Correction *Carolyn Pitchik and Andrew Schotter* 1164

The following Statement of Ownership, Management, and Circulation is provided in accordance with the requirements, as contained in 39 U.S. Code 3685. The *American Economic Review* is owned, managed, and published by the American Economic Association, a nonprofit scientific organization, located at 1313 21st Avenue South, Suite 809, Nashville, Davidson County, Tennessee 37212–2786. The Editor is Professor Orley Ashenfelter, *American Economic Review*, 209 Nassau Street, Princeton, NJ 08542–4607. During the preceding 12 months the average number of copies printed for each issue was 27,900, the average paid circulation, 24,328; the average free distribution, 429; the average number of copies distributed, 24,757. Corresponding figures for the last isssue before filing: 28,000 total number copies printed; 24,956 total paid circulation; 433 free copies distributed; 25,389 total distribution.

An Empirical Study of an Auction with Asymmetric Information

By KENNETH HENDRICKS AND ROBERT H. PORTER*

This paper examines federal auctions for drainage leases on the Outer Continental Shelf from 1959 to 1969. These are leases which are adjacent to tracts on which a deposit has been discovered. We find that the data suggest that neighbor firms are better informed about the value of a lease than non-neighbor firms, that neighbor firms coordinate their bidding decisions, and that both types of firms bid strategically in accordance with the Bayesian-Nash equilibrium.

Since the seminal work of Robert Wilson (1967, 1977), auction theory has recognized that auctions in which information about the value of the object being sold is symmetrically distributed among participants are qualitatively different from those in which information is asymmetrically distributed. In the context of a common value auction, information is said to be asymmetric if the precisions of the signals observed vary across the participants. A polar case is an auction in which one agent has (exact) private information about the value of the object, and others have access only to (noisy) public information. The theoretical work by M. Weverbergh (1979) and Richard Engelbrecht-Wiggans, Paul Milgrom, and Robert Weber (1983) that followed Wilson's articles has focused on this case, and provided a general characterization of Bayesian Nash equilibria. The purpose of this paper is to adapt this theory to the institutional framework of a particular auction, the federal offshore oil and gas drainage lease sales, in which information is asymmetric, and to confront the predictions of this theory with field data. A virtue of our data set is that it is possible to identify agents with superior information, and indeed to quantify the information available to them and to the other, relatively less informed agents.

A drainage sale consists of the simultaneous auction of tracts which are adjacent to tracts on which deposits have been discovered. By contrast, a wildcat sale consists of tracts in areas that have not been drilled, and on which firms are permitted to acquire only seismic information. Bidding and drilling behavior, and *ex post* returns, differ significantly on these two types of tracts. Table 1 gives selected statistics on the sample of wildcat and drainage tracts off the coasts of Louisiana and Texas that were sold by the federal government during the period 1954 to 1969. The figures are based on bidding, drilling, and production data which the federal government provides on each tract, and on the annual survey of drilling costs conducted by the American Petroleum Institute. All dollar figures are in millions of 1972 dollars. Tract value is the estimated *ex post* present value of revenues minus drilling costs. Net profit is estimated tract value minus the discounted royalty and bonus payments made by the winning firm. The numbers in parentheses are standard deviations of the sample means.

Table 1 reveals several striking facts about the sample. The fractions of drainage tracts drilled, and which contained oil, were substantially higher than those of wildcat tracts. The average value of drainage tracts was

---

*Hendricks: Department of Economics, University of British Columbia, B.C. V6T 1Y2. Porter: Department of Economics, Northwestern University, Evanston, IL 60208. This research was funded by the National Science Foundation under grant nos. SES-8511068 and SES-8721231. We are indebted to Harry Paarsch and Richard Spady for advice, and to Michael Riordan for helpful discussions. Much of this research was conducted while Hendricks was visiting the Hoover Institution and Porter was at Bell Communications Research. An earlier version of this paper was presented at the 1987 North American Summer meetings of the Econometric Society at Berkeley, and at a number of universities. We are grateful to the participants in these sessions, and to the referees for their comments.

TABLE 1—SELECTED STATISTICS ON WILDCAT AND DRAINAGE TRACTS[a]

	Wildcat	Drainage
Number of Tracts	1056	144
Number of Tracts Drilled	748	124
Number of Productive Tracts	385	86
Average Winning Bid	2.67	5.76
	(0.18)	(1.07)
Average Net Profits	1.22	4.63
	(0.50)	(1.59)
Average Tract Value	5.27	13.51
	(0.64)	(2.84)
Average Number of Bidders	3.46	2.73

[a] *Source:* Kenneth Hendricks, Robert Porter, and Bryan Boudreau (1987). Dollar figures are in millions of $1972. The numbers in parentheses are standard deviations of the sample means.

more than twice the average value of wildcat tracts. Yet, there was less competition, and profit was roughly four times higher on drainage tracts than on wildcat tracts. The profit differential was even greater when measured in dollars per acre, since drainage tracts were typically half the size of wildcat tracts.[1] The government captured 77 percent of the value of wildcat tracts, but only 66 percent of the value of drainage tracts. Thus, even though drainage tracts were lower risk investments and yielded a significantly higher rate of return, firms were less likely to participate in these auctions. What can explain these facts?

The main difference between wildcat and drainage auctions is the distribution of information. Information in a wildcat auction is essentially symmetric, since the precision of seismic survey information is not likely to vary much across firms. This is not true in drainage auctions. Firms which own neighbor tracts obtain information about the drainage tract from their drilling activities on adjacent tracts. Non-neighbor firms derive their information from private seismic surveys, and observable production on adjacent tracts. The latter sources of information are imperfect substitutes for the information that on-site drilling on adjacent tracts can reveal. Consequently, neighbor firms are likely to be better informed than non-neighbor firms, which, if true, would give them an advantage in bidding against the latter. Non-neighbor firms would have to bid cautiously, if at all, since they would have to worry that their bids will win only if the neighbors' estimate is low. (This affliction is often called the "Winner's Curse.")

We find that the data strongly support this hypothesis. Conditional on publicly available information, the participation and bidding decisions of neighbor firms are significantly better predictors of tract profitability than the participation and bidding decisions of non-neighbor firms. Neighbor firms won most of the profitable drainage tracts, and their average share of the value of drainage tracts is about 44 percent. By contrast, non-neighbor firms earned approximately zero profits.

A naive theory of bidding in a drainage auction with one neighbor firm might predict that non-neighbor firms will not bid, on the grounds that they can never make money against a better-informed neighbor firm. However, such reasoning requires firms to hold incorrect beliefs about the bidding behavior of their rivals: If non-neighbor firms choose not to participate, and the neighbor firm correctly anticipates this strategy, its optimal response is to bid the reservation price when it is worthwhile. But, in that case, the non-neighbor can bid slightly more, win the auction, and earn positive profits on average. Thus, for the firms' behavior to be consistent with an equilibrium model of bidding, non-neighbor firms must behave strategically, and participate in such a manner that the neighbor firm is forced to consider the possibility that it will lose the tract if it bids too low.

We find that the data are consistent with the predictions of the Bayesian Nash equilibrium model of bidding in first-price, sealed bid auction with asymmetric information. Non-neighbor firms were relatively cautious in their bidding, but at least one non-neighbor firm bid in 69 percent of the auctions. The number of non-neighbor bids was more than twice that of neighbor bids, but neighbor firms won well over one-half of the drainage tracts on which they bid. Average net profits to non-neighbor firms were significantly negative on the set of tracts where no neighbor firm bid, and positive on the set of tracts where a neighbor firm bid. As mentioned earlier, average net profits of non-neighbor firms were approximately zero. We also find that the restrictions which equilibrium imposes on the joint distribution of neighbor and non-neighbor bids conditional on tract profitability and publicly available information are not rejected by the data.

---

[1] Walter Mead et al. (1984) obtain similar results, in that the internal rates of return they calculate are higher on drainage tracts. They also note that these returns are higher for firms owning neighbor tracts.

The strong support for the asymmetric information model with one informed bidder was somewhat surprising, since approximately two-thirds of the sample of drainage tracts had multiple neighbor firms. In these cases, competitive bidding among neighbor firms should have eliminated most, if not all, of the information rents. The fact that these rents were positive, and large, suggests that neighbor firms may not have competed against each other. Several facts support this hypothesis. First, there is no law prohibiting firms from forming a bidding consortium in federal offshore auctions, and the neighbor firms may have previously formed such a consortium in order to manage production from the common pool. Second, there were 74 tracts with multiple neighbor firms, but only 17 tracts had multiple neighbor bids. Third, net profits were not significantly lower on tracts with multiple neighbors than on tracts with one neighbor firm. Fourth, the bids of the neighbor firms are strictly decreasing in the number of neighbor firms. Given plausible assumptions on the distributions and given the range of the data, this result is inconsistent with the theory of competitive bidding (see Albert Smiley, 1979). It is consistent with neighbor firms coordinating their bidding decisions, and submitting one serious bid on tracts which are considered worthwhile.

The paper is organized as follows: in Section I, we study an auction in which one firm has private information about the drainage tract, and others have access only to public information. In Section II, the data and estimation strategy are discussed. In Section III, we examine the evidence for the hypotheses that neighbor firms are better informed than non-neighbor firms, and that they bid according to the bidding model described in Section I. In Section IV, we investigate an alternative model in which neighbor firms bid competitively against each other.

I. The Bidding Model

Our bidding model is a version of the noncooperative first-price, sealed bid model with asymmetric information that was introduced by Wilson (1967), and subsequently studied by Weverbergh (1979) and Engelbrecht-Wiggans, Milgrom, and Weber (EMW) (1983). The focus is on the individual tract, and we ignore any structural or strategic factors which may link a firm's bidding decisions on different tracts. Specifically, we assume that (*i*) there are no information externalities between tracts sold in the same sale; (*ii*) each firm is risk neutral; and (*iii*) the bidding strategy of each firm for a tract depends only on the state of information and competition for that tract. Assumption (*i*) is justified by the observation that drainage tracts in a sale are usually drawn from geographically distinct areas. Assumption (*ii*) is not unreasonable since, for most of the participants, bids in an OCS drainage sale represent only a small part of their annual exploration budgets. Assumption (*iii*) implies competitive bidding between the neighbor and non-neighbor firms. It rules out the use of punishment strategies in which bidding behavior on a tract is made contingent on bidding outcomes on tracts sold in previous sales.

There is one neighbor firm and an arbitrary number of non-neighbor firms. Let X and Z denote, respectively, the private and public signals on V, the unknown value of the representative drainage tract. The neighbor firm observes the realizations of X and Z prior to bidding on the tract, while non-neighbor firms observe only the realization of Z. Realizations of the random variables will be denoted by lowercase letters. In what follows, we treat z as given, and are explicit about the dependence of the distributions on its value. However, for notational convenience, we will suppress the dependence of bidding strategies on z.

The essential feature of our model is that the information revealed by on-site drilling of an adjacent tract by a neighbor firm is a sufficient statistic for the information non-neighbor firms acquire from seismic surveys. The assumption that this information is known to the neighbor firm is made in order to obtain a precise characterization of the equilibrium and its properties. A more realistic, but less tractable, assumption is that the non-neighbor firms have noisy, but private, estimates of tract value.[2] However, as long as the estimates of the non-neighbor firms are not too informative, we can use the result by Milgrom and Weber (1985) on the upper hemicontinuity of the equilibrium correspondence to argue that the behavioral implications of this descriptively more accurate model are approximately the same as those of a model in which the estimates of non-neighbor firms are based on public information.

The strategy of non-neighbor firm i is a distribution function $G_i(\cdot)$ over the nonnegative real numbers. Adopting the approach of EMW, we summarize the information of the neighbor firm by the real-valued, random variable $H = E[V|X, z]$. We shall assume that H has an atomless distribution, $F(\cdot|z)$, with finite mean, $\bar{H}$. The strategy of the neighbor firm can then be defined as a function σ which maps realizations of H, which are associated with the realizations of X, into the nonnegative real numbers. We shall assume that $\sigma(h)$ is a differentiable, strictly increasing function on the range (R, ∞), where R is the reservation price, and denote its inverse function on this interval by $\tau(b)$.

Define $G(b) = G_1(b) \cdots G_n(b)$ to be the distribution function of the maximum of the bids submitted by the uninformed firms on the tract. Given the strategy combination $(\sigma, G_1, \ldots, G_n)$, the payoff to the neighbor firm, when its estimate of V is h, is the product of the probability that its bid is highest and its expected value of the tract less its bid.

$$G(\sigma(h))(h-\sigma(h)). \tag{1}$$

If the drainage tract contains any oil, it is usually part of a pool which the neighbor firm has discovered on the adjacent tract. This makes the value of the drainage tract to each firm contingent upon the manner in which production is allocated among the firms. If the firms bargain to an efficient allocation, tract valuations are identical across firms. In many instances, however, competition leads to some dissipation of rents (see Gary Libecap and Steven Wiggins, 1985). In these cases, the neighbor firm is likely to have a higher tract valuation than the non-neighbor firm, since it can take the externality into account and internalize its effects.

We parameterize the possible difference in tract valuations by letting the expected value of the drainage tract to the non-neighbor firm be equal to $E[H|z] - c$, where c is a fixed, nonnegative constant. (This constant could also reflect any cost differences.) The expected payoff to non-neighbor firm i which submits a bid b greater than R is

$$E[H - b - c|\tau(b) > h; z] \cdot F(\tau(b)|z) \cdot \Pi_{j \neq i} G_j(b). \tag{2}$$

The first term in equation (2) is expected profits, conditional on winning the tract (and hence $b > \sigma(h)$). The remaining terms represent the probabilities of outbidding the neighboring firm and the other non-neighbors. Ties at the reservation price are assumed to be settled by randomization.

A Bayesian Nash equilibrium for the bidding game is an $(n+1)$-tuple of strategies $(\sigma^*, G_1^*, \ldots, G_n^*)$ such that the expected payoff to each firm conditional on its information is maximized, given the strategies employed by the other firms.

---

[2] See Wilson (1975) for an analysis of an example of such an auction.

We turn next to a characterization of the equilibrium bid distributions. Define

$$(3)\quad \phi(h) = \exp\left\{ -\int_h^\infty \frac{f(s;z)\int_h^s F(u;z)\,du\,ds}{cF(s;z)^2 + F(s;z)\int_h^s F(u;z)\,du} \right\}.$$

Note that if c is equal to zero, then $\phi(h)$ is equal to $F(h;z)$. Our theorem is a restatement of the theorem proved by EMW, extended to auctions with asymmetric tract valuations. The proof is essentially the same as the one given by EMW, and is given in the Appendix.

THEOREM: *The* $(n+1)$*-tuple* $(\sigma^*, G_1^*, \ldots, G_n^*)$ *is an equilibrium point if and only if*

$$G^*(b) = \begin{cases} 1 & b > \bar{H} - c \\ \phi(\tau(b)) & R < b < \bar{H} - c \\ \phi(R) & 0 \le b < R \end{cases}$$

$$\sigma^*(h) = \begin{cases} E[H \mid H \le h; z] - c & h > \hat{h} \\ R & \hat{h} \ge h \ge R \\ 0 & h < R \end{cases}$$

where $\hat{h}$ *solves* $E[H \mid H < \hat{h}; z] - c = R$.

The theorem states that the supports of the equilibrium distribution functions are identical, and consist of $\{0\}$ and the interval $[R, \bar{H} - c]$. We interpret a zero bid as no bid. The equilibrium strategy of the neighbor firm on $(R, \bar{H} - c]$ is uniquely determined by the condition that, in equilibrium, non-neighbor firms must earn zero profits. That is, suppose a non-neighbor firm submits an equilibrium bid b. Then, since σ^* is strictly increasing at b, there is a unique h' such that $b = \sigma^*(h')$. The expected profit of the tract to the non-neighbor firm conditional on the event that it wins is $E[H \mid H < h'; z] - b - c$. Setting this equation equal to zero implies that $\sigma^*(h') = E[H \mid H < h'; z] - c$.

The equilibrium strategies of non-neighbor firms are indeterminate. However, the equilibrium distribution function of the maximum uninformed bid is unique. It is chosen in order to induce the neighbor firm to bid according to the function given above. The two distributions differ in the probability of the events of no bid, and of a reservation bid R. G^* possesses a mass point equal to $F(\hat{h})$ at $\{0\}$, and is constant at this value on the interval $(0, R]$. The distribution of the neighbor bid also possesses a mass point at $\{0\}$, but it is equal to $F(R)$, which is less than $F(\hat{h})$. The distribution is constant at $F(R)$ on the open interval $(0, R)$, and then jumps discontinuously upward at R. The value of the mass point at R is equal to $F(\hat{h}) - F(R)$. If c is equal to zero, G^* is identical to the distribution of the informed bid on $(R, \bar{H})$.

The randomized strategies of non-neighbor firms are a direct consequence of the assumption that the neighbor firm knows their estimates. If non-neighbor firms bid according to a pure (and hence predictable) strategy which specifies a bid for each realization of the public information variables, the optimal response of the neighbor firm is to bid slightly above the maximum non-neighbor bid if the tract is worth more than this number, and not bid otherwise. But this implies that on average the non-neighbor firm is certain to lose, since it will win only those tracts whose expected value is less than its bids. By randomizing, non-neighbor firms can induce the neighbor firm to bid according to a strategy in which it will lose profitable tracts some of the time. As a result, non-neighbor firms will earn positive expected profits on some tracts, and it is only on average that their expected profits are zero.

Nevertheless, some readers may find the mixed strategy equilibrium objectionable, since non-neighbor firms obviously do not determine their bids by spinning roulette wheels. Mixed strategy equilibria, however, can often be justified as a limit of a sequence of pure strategy equilibria in an appropriately specified perturbed game. In our context, a natural candidate perturbed game is one in which the signals of non-neighbor firms are noisy estimates of V, and are independently and continuously distributed conditional on X, the information of the neighbor firm. The theorems of Milgrom and

Weber (1985) imply that there is an equilibrium in this game where each firm's bid is a function of its private estimate, and which distributionally approximates the mixed strategy equilibria of our model.

A. *Properties of Equilibrium*

The public information variable Z affects the bidding strategy of the neighbor firm in two ways, through its expectations about V and its response to its rivals. Integrating by parts, σ can be expressed as

$$(4)\quad \sigma(h;z) = h - c - \int_{-\infty}^{h} F(s;z)\,ds/F(h;z).$$

The integral term represents the factor by which the neighbor shades its bid downward from its estimate of the value of tract. Equation (4) reveals that even if, given the private signal, the public signal is redundant information (i.e., $E[V|X,Z]=E[V|X]$), the public signal continues to play an important strategic role in determining the neighbor firm's bid. It affects the magnitude of the "shading" factor. Consequently, any public information variable which helps the non-neighbor firms to predict the value of the drainage tract will be an important explanatory variable in the bidding equation of the neighbor firm.

Under what conditions is this relationship monotone? Given any h in the support of F, define the probability distribution function

$$k(H;z) = \begin{cases} f(H;z)/F(h;z) & H<h \\ 1 & H\geq h. \end{cases}$$

It is easily verified that $k(\cdot\,;z)$ has the monotone likelihood ratio property (MLRP) if and only if f has this property. It then follows from the well-known result that distribution functions which possess the MLRP can be ranked by first-order stochastic dominance that $E[H|H<h,z]$ is an increasing function of z. Using this result and equation (4), one can then show that, conditional on its private signal, the neighbor firm bids higher when non-neighbor firms are more optimistic about the value of the drainage tract. In what follows, we will assume that f has the MLRP.

The *ex ante* probability of the neighbor firm winning a drainage tract, conditional on the event that at least one firm bids, is easily calculated when c is equal to zero. Recall that, in this case, $G^*(\sigma(h)) = F(h)$ for all $\sigma(h) > R$. Integrating $F(h)$ over h using the density f, and taking into account the mass points at $\{0\}$ and $\{R\}$, yields the expression for the *ex ante* probability of the neighbor firm winning,

$$1/2 + (1/2)F(\hat{h};z)\left[F(\hat{h};z) - F(R;z)\right] \Big/ \left[1 - F(\hat{h};z)F(R;z)\right],$$

which is strictly greater than 1/2. This implies that the neighbor firm is likely to win more than one-half of the drainage tracts in our sample.

An increase in c causes the neighbor firm to bid less at every value of h, and, as a result, the cutoff value $\hat{h}$ rises. This increases the likelihood of the event that the neighbor firm bids R, and of the event that no non-neighbor firm bids are submitted. It can also be shown that the likelihood of the event that the neighbor firm wins the drainage tract is increasing in c.

The number of non-neighbor firms has no effect on the equilibrium bid distributions and payoffs, provided this number is not zero. From an empirical viewpoint, this is a particularly desirable feature of the model. The number of non-neighbor firms is not an observable variable to the econometrician. We observe only the number of non-neighbor firms which choose to participate in the auction. In most bidding models, this would lead to a selection bias in the estimates of the coefficients of the bid functions. However, in our context, the problem does not arise.

In summary, we list below the main empirical predictions of the model.

1. The event that no neighbor firm bids occurs less frequently than the event that no non-neighbor firm bids.
2. The neighbor firm wins at least one-half of the tracts.

3. Expected profits to non-neighbor firms are zero. They are negative on the set of tracts where no neighbor firm bids, and positive on the set of tracts where the neighbor firm bids.

4. Expected profits to the neighbor firm incorporates an information premium which makes its earnings above "average."

5. If c is equal to zero, the *ex ante* bid distributions (i.e., prior to the realization of X) are approximately the same.

6. The bidding strategy of the neighbor firm is independent of the number of non-neighbor firms.

7. The bidding strategy of the neighbor firm is an increasing function of the public signal, when a larger signal is "good news."

II. Data and Estimation Methods

Our study focuses on the federal lands off the coasts of Louisiana and Texas which were leased between 1959 and 1969. During this period, the government held 8 drainage sales, in which it auctioned off 144 tracts. This number does not include 25 drainage tracts on which the high bid was rejected, for which we have no data.

In our sample, each lease is sold via a first-price, sealed bid auction. A bid is a dollar figure which the firm promises to pay to the government at the time of the sale if it is awarded the tract. This payment is called the *bonus*. The terms of the lease are that, if no exploratory work is done after five years have elapsed, then ownership of the lease reverts to the government. If oil and/or gas is discovered in sufficient quantities so that the firm begins production, the lease is automatically renewed for as long as it takes the firm to extract the hydrocarbons. A fixed fraction of the revenues from any oil and/or gas extracted, one-sixth throughout our sample, accrues to the government. This sum is paid on an annual basis and is called the *royalty* payment. A nominal rental fee ($3 per acre on wildcat tracts, and $10 per acre on drainage tracts) is paid by the firm each year until either the lease expires or production begins.

The government may enter the auction as a bidder in two ways. In our sample, it announced a reservation price of $25 per acre on most drainage leases. (The reservation prices varied from sale to sale.) In addition, it retains the right to reject the high bid on a tract if it believes the bid is too low. The usual basis on which it makes this judgment is its private estimate of the value of the tract. These estimates may be based in part upon the geological and seismic reports which the firms are required to submit. For sales in our sample, the high bid was rejected on 7 percent of the wildcat tracts, and on 15 percent of the drainage tracts.

Our data set contains the following information for each tract: the date it was sold; its location and acreage; which firms bid and the value of their bids; the number and date of any wells that were drilled; and annual production through 1980 if any oil or gas was extracted. The drilling and production data were used, together with the annual survey of drilling costs conducted by the American Petroleum Institute, to calculate *ex post* discounted revenues and costs for each tract. Real wellhead prices in the United States were virtually constant from 1950 until 1973, and we assume that the expectations of the bidders in our sample would be that this pattern would continue. Accordingly, future production paths were converted into revenues by using the real wellhead prices at the date of sale, and discounted to the auction date at a 5 percent per annum rate. See our previous paper with Boudreau for further detail.

From the original sample of 144 drainage tracts, we selected 114 tracts which were adjacent to previously leased federal tracts. (The remaining tracts were adjacent to state tracts, about which we have no information.) For each drainage tract, we then designated neighboring tracts as those which had previously been sold and were adjacent to it, and designated firms as neighbors if they had purchased the rights to one of these tracts. By the same methods we used for the drainage tracts, we computed discounted revenues and costs for each of the neighboring tracts.

The tracts are typically in a square grid pattern, but can vary in size. Wildcat tracts are usually either 5,000 or 5,760 acres,

	0	1	2	3	4	5	6
No. of tracts	0	40	43	21	5	4	1

FIGURE 1. NUMBER OF NEIGHBOR FIRMS

	0	1	2	3	4	5	6
No. of tracts	19	79	15	0	0	1	0

FIGURE 2. NUMBER OF NEIGHBOR BIDS

TABLE 2—DEFINITION OF VARIABLES[a]

	Mean	Standard Deviation
B_I: maximum bid by neighbor	3.78	11.52
B_U: maximum bid by non-neighbor	3.60	9.57
N_I: number of neighbor bids	1.00	0.67
N_U: number of non-neighbor bids	1.69	2.09
N: number of neighbor tracts	3.01	1.98
NF: number of neighbor firms	2.06	1.08
π: *ex post* tract gross profitability	8.75	20.83
V: *ex post* gross profits of adjacent tract	14.51	20.16
A: tract acreage	2.679	1.533

[a] Dollar figures are in millions of $1972. Tract acreage is in thousands of acres.

and drainage tracts are often 2,500 acres or less. Consequently, the number of possible neighbor tracts is never less than eight, and is sometimes larger. The actual number is usually much less, and the number of neighbor firms is even smaller, since one firm frequently owned more than one neighbor tract. The frequency distribution of neighbor firms per drainage tract is given in Figure 1.

There were 74 tracts with more than one neighbor firm. However, in most of these cases, only one of the neighbor firms bid. The frequency distribution for the number of neighbor bids per drainage tract is given in Figure 2.

The fact that only 16 of the 74 tracts received more than one neighbor bid suggests that neighbor firms may have coordinated their participation and bidding decisions. Such behavior was not prohibited by the federal government in offshore oil auctions during the sample period. In fact, neighbor firms may have formed a joint venture prior to the sale in order to manage production from the common pool. This would also have provided neighbor firms with a mechanism for distributing the benefits from cooperation. Neighbor firms which did not bid could have received transfer payments through the allocation of production shares.

In what follows, we shall assume that the neighbor firms coordinated their bid decisions and submitted one serious bid on tracts which were considered worthwhile. The alternative hypothesis of competitive bidding among neighbor firms is examined in Section IV.

Table 2 lists the empirical analogues of the theoretical variables. The largest neighbor bid is denoted by B_I, or the reservation price in the event that no neighbor firm bid. Similarly, the largest non-neighbor bid is denoted by B_U, or the reservation price in the event that no non-neighbor firm bid. The number of non-neighbor bids is given by N_U, the number of neighbor bids by N_I, and the number of neighbor firms by NF. Our proxy for the private information of the neighbor firms is the gross profitability of the tract, which is denoted by π. This is defined as discounted revenues less drilling costs and less royalty payments (one-sixth of revenues in our sample). It is *not* net of the bid. Our proxies for the public information variables are: the number of neighbor tracts (N), the gross profitability of the most recently sold neighbor tract, or the average of these values if more than one neighbor tract was sold at the same time (V), and tract acreage (A).

A. Likelihood Function

The estimates reported in Tables 6, 7, and 8 are obtained from similar likelihood functions. Define indicator variables for the maximum neighbor and non-neighbor bids:

$$Y_{it} = W_{it}'\theta_i + \varepsilon_{it}, \qquad i = I, U;\ t = 1, \dots, T$$

where W_{it} is a vector of regressors for tract t, θ_i is a parameter vector, and $\{\varepsilon_{It}, \varepsilon_{Ut}\}$ are i.i.d. drawings from a bivariate normal distribution with zero mean, variances σ_I^2 and σ_U^2, and covariance σ_{IU}. Here T is the number of tracts which received at least one bid. Bids are determined as follows:

$$\begin{aligned} \log(B_{it}/R_t) &= Y_{it} \quad \text{if } Y_{it} \geq 0 \\ &= 0 \quad \text{otherwise,} \end{aligned}$$

where R_t is the reservation price on tract t. Thus, bids are assumed to be lognormally distributed, in accord with previous evidence. (See Smiley, 1979, for example.) We have data only for tracts in which at least one bid is positive. Therefore, we are faced with both truncated dependent variables and sample selection issues.

We partitioned the tracts into three mutually exclusive sets: Ω_{++} is the set of tracts with at least one neighbor bid and at least one non-neighbor bid, Ω_{+0} is the set of tracts with at least one neighbor bid and no non-neighbor bids, and Ω_{0+} is the set of tracts with no neighbor bids and at least one non-neighbor bid. The log likelihood function for the entire sample can then be defined as follows:

$$\log L = \sum_{t \in \Omega_{++}} l_{1t} + \sum_{t \in \Omega_{+0}} l_{2t} + \sum_{t \in \Omega_{0+}} l_{3t}$$

where

$$\begin{aligned} l_{1t} = &-\left[\log(2\pi) + (1/2)\log|\Sigma|\right] \\ &-(1/2)(\varepsilon_{It}, \varepsilon_{Ut})\Sigma^{-1}(\varepsilon_{It}, \varepsilon_{Ut})' \\ &-\log(1 - Z(-W_{Ut}\theta_U/\sigma_U, \\ &\qquad -W_{It}\theta_I/\sigma_I; \rho_{IU})), \end{aligned}$$

$$\begin{aligned} l_{2t} = &\log\Big(\Phi\Big\{\left[-W_{Ut}\theta_U - \sigma_{IU}\sigma_I^{-2}\varepsilon_{It}\right] \\ &\qquad /\left[\sigma_U^2 - \sigma_{IU}^2\sigma_I^{-2}\right]^{1/2}\Big\}\Big) \\ &-\log(\sigma_I) + \log(\phi(\varepsilon_{It}/\sigma_I)) \\ &-\log(1 - Z(-W_{Ut}\theta_U/\sigma_U, \\ &\qquad -W_{It}\theta_I/\sigma_I; \rho_{IU})), \end{aligned}$$

$$\begin{aligned} l_{3t} = &\log\Big(\Phi\Big\{\left[-W_{It}\theta_I - \sigma_{IU}\sigma_U^{-2}\varepsilon_{Ut}\right] \\ &\qquad /\left[\sigma_I^2 - \sigma_{IU}^2\sigma_U^{-2}\right]^{1/2}\Big\}\Big) \\ &-\log(\sigma_U) + \log(\phi(\varepsilon_{Ut}/\sigma_U)) \\ &-\log(1 - Z(-W_{Ut}\theta_U/\sigma_U, \\ &\qquad -W_{It}\theta_I/\sigma_I; \rho_{IU})). \end{aligned}$$

Here Σ is the covariance matrix, ρ_{IU} is the correlation coefficient, Φ and ϕ are the standard normal distribution function and density function, respectively, and $Z(\cdot, \cdot)$ is the function for the probability of the event that no bid is observed on a tract, as defined in Norman Johnson and Samuel Kotz (1972, p. 93).

III. The Evidence

We begin by examining the predictions of the model with respect to the participation rates, and returns of neighbor and non-

TABLE 3—SAMPLE STATISTICS ON TRACTS WON BY EACH TYPE OF FIRM[a]

	Wins by Neighbor Firms		Wins by Non-Neighbor Firms		
	A	Total	B	C	Total
No. of Tracts	35	59	19	36	55
No. of Tracts Drilled	23	47	18	33	51
No. of Productive Tracts	16	36	12	19	31
Average Winning Bid	3.28	6.04	2.15	6.30	4.87
	(0.56)	(2.00)	(0.67)	(1.31)	(0.92)
Average Gross Profits	10.05	12.75	−0.54	7.08	4.45
	(3.91)	(3.21)	(0.47)	(2.95)	(1.99)
Average Net Profits	6.76	6.71	−2.69	0.78	−0.42
	(3.02)	(2.69)	(0.86)	(2.64)	(1.76)

[a] Dollar figures are in millions of $1972. The numbers in parentheses are the standard deviations of the sample means. Column A refers to tracts which received no non-neighbor firm bid, column B refers to tracts which received no neighbor bid, and column C to those in which a neighbor firm bid, but a non-neighbor firm won the tract.

neighbor firms. The number of tracts which received no neighbor bid is 19, and the number of tracts which received no non-neighbor bid is 35. Therefore, at least one neighbor firm participated in 83 percent of the auctions, and at least one non-neighbor firm participated in 68 percent of the auctions. This is consistent with the theoretical model.

Table 3 gives sample statistics on the tracts won by each type of firm. Column A refers to tracts which received no non-neighbor firm bid, column B refers to tracts which received no neighbor bid, and column C to those in which a neighbor firm bid, but a non-neighbor firm won the tract.

The evidence is consistent with the model. The neighbor firm won 62 percent of the tracts that it bid on. As we calculated in our previous paper, its share of the tract value was approximately 44 percent, which was considerably higher than the 23 percent average firm share on wildcat tracts. The average net profit of non-neighbor winners was virtually zero. It was positive on tracts which received a neighbor bid, and it was significantly negative on tracts which received no neighbor bid. By contrast, the participation decisions of the non-neighbor firms had no effect on the earnings of neighbor firms. Based on these return figures, it appears as if the neighbor firm was better able to identify which drainage tracts were more likely to contain oil, and was able to exploit this knowledge to obtain above average profits.

We found no evidence of a mass point at the reservation price in the distribution of bids of the neighbor firm. One possible explanation for this result is that firms were afraid that the government would reject reservation price bids. Recall that the government rejected the high bid on 25 drainage tracts.[3]

The existence of a positive reservation price provides an explanation for why non-neighbor firms drilled tracts on which no neighbor firm bid. The lack of participation by the neighbor firm implies that its expectation of net profit is less than R. Since R is positive, and the bid is a sunk cost, it may still have been rational for the non-neighbor firm to drill its lease. Drilling outcomes were not inconsistent with this belief, since the average gross profit of tracts which received no neighbor bid was not significantly different from zero at conventional confidence levels.

An indirect test of our assumption that neighbor firms coordinated their bidding decisions is to compare the bidding behavior of neighbor firms and their net profits on single neighbor tracts to that on multiple neighbor

---

[3] We study the issue of a random reservation price, and its empirical implications, in a subsequent paper.

TABLE 4—THE EFFECT OF NEIGHBOR FIRM COMPETITION ON NEIGHBOR FIRM PARTICIPATION AND PROFITS[a]

	Single Neighbor Tracts	Multiple Neighbor Tracts No. of Neighbor Bids		
		1	≥ 2	Total
No. of Tracts	40	48	15	74
No. of Tracts with No Neighbor Bid	8	–	–	11
No. of Wins	19	29	11	40
Average Winning Bid of Neighbor Firm	4.795 (1.444)	2.615 (0.697)	17.193 (9.953)	6.624 (2.885)
Average Gross Profits of Neighbor Firm	13.601 (5.608)	4.670 (2.148)	32.597 (11.506)	12.350 (3.965)
Average Net Profits of Neighbor Firm	8.806 (4.762)	2.055 (1.690)	15.404 (10.963)	5.725 (3.297)

[a] Dollar figures are in millions of $1972. The numbers in parentheses are the standard deviations of the sample means.

tracts. If our assumption is correct, they should not be significantly different across the two categories.

The statistics reported in Table 4 are consistent with this prediction. In each category, the neighbor firm won approximately one-half of the tracts. The net profits on the single neighbor tracts won by neighbor firms is somewhat higher than on the multiple neighbor tracts won by neighbor firms, but the difference is not statistically significant. In both categories, net profits were significantly positive, and quite large. High value multiple neighbor tracts tended to attract more than one neighbor bid, but average net profits were substantially *higher* on these tracts than on the multiple neighbor tracts with one neighbor bid. This may be an indication of "shadow" bidding, in which the neighbor firms submit more than one bid in order to convince the government that bidding is competitive.

A. *Information Structure*

An important assumption of our theoretical model, both in terms of its predictive consequences and its influence on our empirical formulation, is that non-neighbor firms have access only to publicly available information. We assume that they have no private signals which are useful in predicting tract profitability, and are not also observed by the neighbor firm.

While we cannot observe the firms' private information signals, it is possible to test this assumption indirectly. In particular, we conducted the following predictive exercise. We first regressed the gross profitability of a given drainage tract on indices of whether a neighbor firm bid (the dummy variable $D_I = 1$ if so), on the number of neighbor bids, on tract acreage, on the number of neighbor tracts, and on a second-order polynomial in the maximum neighbor firm bid and the value of the neighboring tract. We then supplemented this set of regressors with an index of whether any non-neighbor firms bid ($D_U = 1$ in these cases) and, if so, how many do so. We also included the maximum uninformed bid as an additional variable in the second-order polynomial expression. We then performed the same exercise with the roles of the neighbor and non-neighbor reversed. If non-neighbor firms had access to informative private signals, then one would expect their participation and bidding decisions to have some predictive content. If not, then the assumption that no payoff-relevant private information was available to non-neighbor firms may be correct. We have included public information such as acreage, number of neighbor tracts, and adjacent tract value, since the relationship between profitability and bids depend on these values in equilibrium.

In Table 5 we report three regression equations, corresponding to the three sets of

TABLE 5—PREDICTION OF TRACT PROFITABILITY[a]

Variable	Equation (1)	Equation (2)	Equation (3)
Constant	-3.60	0.11	3.16
	(-0.75)	(0.02)	(0.81)
D_I	5.03	-3.12	
	(1.42)	(-0.60)	
D_U		-0.09	4.89
		(-0.03)	(1.03)
N_I	3.98	-2.01	
	(1.01)	(-0.46)	
N_U		1.64	0.93
		(1.68)	(0.87)
N	-0.26	0.003	-1.03
	(-0.48)	(0.007)	(-1.41)
A	$-.46$	-0.18	0.80
	(-0.62)	(-0.23)	(0.67)
B_I	3.55	3.09	
	(3.90)	(3.35)	
B_I^2	-0.023	0.061	
	(-2.57)	(1.76)	
B_U		-0.229	0.181
		(-0.31)	(0.26)
B_U^2		0.014	0.021
		(0.65)	(1.36)
V	0.013	0.116	0.259
	(0.11)	(1.30)	(1.80)
V^2	$-0.9E-4$	$-0.41E-3$	-0.0013
	(-0.12)	(-0.37)	(-0.72)
$B_I \cdot B_U$		-0.093	
		(-1.99)	
$B_I \cdot V$	-0.0067	-0.026	
	(-0.24)	(-1.04)	
$B_U \cdot V$		-0.034	-0.034
		(-1.41)	(-0.85)
SSE	17673	14766	41483
R^2	.640	.699	.154
d.o.f.	104	98	104

[a] The dependent variable in each equation is π. Heteroskedasticity-consistent t-statistics are displayed in brackets.

regressors discussed in the preceding paragraph. An F-statistic for the test that none of the non-neighbor firm participation and bid variables coefficients are significantly different from zero, which compares the regressions in columns one and two, equals 3.22. Under the null hypothesis, the statistic has (6,98) degrees of freedom, the critical value for which is 3.71 at size 0.05. Most of the explanatory power of the non-neighbor firm variables is derived from the product term involving the neighbor bid.

We shall henceforward frequently refer to non-neighbor firms as uninformed and neighbor firms as informed. The evidence summarized in Table 5 does not contradict this nomenclature.

Note that the significant coefficients in Table 5 also support the view that the informed firms do indeed possess payoff-relevant information. True tract profitability is positively correlated with their bids, over the entire observed range of bids. The final column indicates that the incremental predictive power of the informed firm bid and participation decisions is very significant, even after conditioning on public information and non-neighbor bid and participation information.

B. *Bid Distributions*

One implication of the theoretical model is that, conditioning solely on publicly available information, the distribution of the informed bid and that of the maximum uninformed bid should be approximately the same if tract valuations are symmetric (i.e., c is equal to zero). Accordingly, we computed the maximum likelihood estimates of the parameters of the joint distribution of these two variables. We explicitly accounted for the truncation of the bid variables at the reservation price, for the sample selection rule that the only observed tracts are those in which at least one bid was positive, and for the possibility of correlation between the error terms of the two bid equations. The explanatory variables are the publicly available information in our sample: tract acreage, the number of neighbor tracts, the value of the adjacent tract, and that value squared.

The maximum likelihood estimates, which are contained in Table 6, have two notable features. First, the coefficients and estimated standard errors of the regression equations are similar. The value of the χ^2 statistic of the null hypothesis that the two regression equations are identical is 10.58, which is below 11.07, the critical value of a χ^2 statistic with 5 degrees of freedom at size 0.05. This result accords weakly with the theoretical prediction of our model when tract valuations are symmetric. Second, there is essentially no correlation between the disturbances of the two equations. The esti-

TABLE 6—JOINT DISTRIBUTION OF BIDS CONDITIONAL ON PUBLIC INFORMATION[a]

	Unrestricted		Restricted
Independent Variable	Dependent Variable $\log(B_I/R)$	$\log(B_U/R)$	Dependent Variable $\log(\mathrm{BID}/R)$
Constant	1.98068	2.05437	−1.99365
	(3.44)	(2.70)	(3.96)
V	0.07391	0.00523	0.04966
	(3.42)	(0.19)	(2.52)
V^2	−0.00073	−0.00009	−0.00050
	(−2.92)	(−0.30)	(−2.17)
A	−0.11092	0.13285	−0.02499
	(−0.82)	(0.74)	(−0.21)
N	−0.08226	−0.28903	−0.14763
	(−0.74)	(−1.97)	(−1.51)
$\begin{bmatrix}\sigma_I & \\ \rho_{IU} & \sigma_U\end{bmatrix}$	$\begin{bmatrix}2.0151 & \\ (11.7) & \\ 0.1034 & 2.6596 \\ (0.94) & (12.7)\end{bmatrix}$		$\begin{bmatrix}2.0528 & \\ (11.5) & \\ 0.0638 & 2.6785 \\ (0.57) & (12.8)\end{bmatrix}$
	$\text{Log } L = -428.895$		$\text{Log } L = -434.184$

[a]Asymptotic t-statistics are in parentheses. They are computed from the analytic second derivatives. They are not appreciably different from the Eicker-White t-statistics.

mated correlation coefficient, ρ_{IU}, is 0.10 and not significantly different from zero. This suggests that there are no omitted variables which might significantly affect both bidding equations. In particular, although tract profitability is correlated with the informed firm's private signals, and so also with the informed firm bid, it appears that this additional information is only weakly correlated with the unexplained component of the maximum uninformed bid.

The theoretical model also predicts that, conditioning on public information and tract profitability, the distributions of the informed bid and the maximum uninformed bid should differ. More precisely, tract profitability should be highly correlated with the neighbor bid, and orthogonal to the maximum non-neighbor bid. To examine this implication of asymmetric information, we computed the maximum likelihood estimates of the joint distribution of bids conditional on tract gross profits and that variable squared, and the public information variables listed above.

The estimation results, which are reported in Table 7, are consistent with the above predictions. The χ^2 statistic of the null hypothesis that the two distributions are the same is 19.84, which is rejected at size 0.01. The coefficients for the tract profitability variables are highly significant in the equation for the neighbor firm bid, and insignificant in the equation for the maximum non-neighbor firm bid. Non-neighbor firms do not appear to have access to information, other than the number and value of neighbor tracts and tract acreage, which is correlated with tract profitability.

The estimates presented in Table 8 can be given a more structural interpretation: they are the maximum likelihood estimates of the coefficients of the informed firm's bid equation and the maximum uninformed firm bid equation, accounting for the truncation of these two variables and the sample selection rule. The maximum uninformed bid is taken to be a function of publicly available information: acreage, number of neighbor tracts, adjacent tract value, and that value squared. The informed bid is a function of these variables, together with our proxies for its private information: actual tract profitability and that figure squared. We also include the number of uninformed firms in both equations. According to the theory, this number

TABLE 7—JOINT DISTRIBUTION OF BIDS CONDITIONAL ON PUBLIC INFORMATION AND TRACT PROFITABILITY[a]

	Unrestricted		Restricted
Independent Variable	Dependent Variable $\log(B_I/R)$	$\log(B_U/R)$	Dependent Variable $\log(\text{BID}/R)$
Constant	1.86237	2.07435	1.88962
	(4.15)	(2.42)	(4.63)
π	0.09102	0.02765	0.07532
	(4.30)	(0.79)	(3.93)
π^2	−0.00053	−0.00026	−0.00046
	(−2.12)	(−0.62)	(−2.09)
V	0.04428	−0.00323	0.03361
	(2.55)	(−0.12)	(2.05)
V^2	−0.00045	−0.00001	−0.00036
	(−2.25)	(−0.03)	(−1.80)
A	−0.20962	0.10221	−0.13419
	(−1.95)	(0.58)	(−1.34)
N	−0.00888	−0.25858	−0.06645
	(−0.10)	(−1.81)	(−0.83)
$\begin{bmatrix}\sigma_I & \\ \rho_{IU} & \sigma_U\end{bmatrix}$	$\begin{bmatrix}1.5996 & \\ (11.5) & \\ 0.0492 & 2.6162 \\ (0.46) & (13.1)\end{bmatrix}$		$\begin{bmatrix}1.6379 & \\ (11.3) & \\ -0.0216 & 2.8014 \\ (-0.20) & (11.8)\end{bmatrix}$
	$\log L = -409.0028$		$\log L = -418.9243$

[a] Asymptotic t-statistics are in parentheses. They are computed from the analytic second derivatives.

should have no explanatory power in the informed bid equation, unless it serves as a proxy for omitted public information variables which might affect uninformed firm bidding. Since the maximum uninformed bid is an order statistic whose distribution depends on N_U, it should be significant in this equation. However, it is properly viewed as endogenous and its coefficient has no structural interpretation.

The estimates in Table 8 indicate that the informed bid is an increasing function of tract profitability and the value of the adjacent tract over the range of values encountered in our sample, and it is essentially independent of the number of neighbor tracts and the number of uninformed bids. The lack of correlation between the maximum informed bid and the number of uninformed bids provides further evidence that our list of public information variables is adequate. If it were not, then unproxied public information (omitted elements of Z) would influence both N_U and B_I.

By contrast, the maximum uninformed bid is not significantly correlated with tract profitability or the value of the adjacent tract, and it is a decreasing function of the number of neighbor tracts. As expected, there is a strong positive correlation between the maximum uninformed bid and the number of uninformed bids.

The sign and magnitudes of the coefficients for the number of neighbor tracts variable in the bid equations are consistent with the maintained hypothesis that neighbor firms do not compete against each other. Under this assumption, the number of neighbor tracts is a proxy for the amount of information which neighbor firms possess. Therefore, non-neighbor firms should bid less aggressively on tracts with a larger number of neighbor tracts, and, in response, the neighbor firm should shade its bid downward.

The signs of the coefficients for the other two public information variables that possess some incremental explanatory power,

TABLE 8—BID EQUATIONS[a]

	Equation (1)		Equation (2)		Equation (3)	
	Dependent Variable		Dependent Variable		Dependent Variable	
Independent Variable	$\log(B_I/R)$	$\log(B_U/R)$	$\log(B_I/R)$	$\log(B_U/R)$	$\log(B_I/R)$	$\log(B_U/R)$
Constant	1.86973	2.13073	1.64933	2.15018	1.67785	0.064395
	(−4.19)	(2.90)	(3.52)	(2.96)	(3.66)	(1.14)
π	0.08967		0.08505		0.08501	
	(4.26)		(4.09)		(4.08)	
π^2	−0.00051		−0.00047		−0.00047	
	(−2.04)		(−1.88)		(−1.88)	
V	0.04452	0.00257	0.04814	0.00120	0.04757	0.02083
	(2.58)	(0.10)	(2.82)	(0.04)	(2.79)	(1.08)
V^2	−0.00045	−0.00006	−0.00047	−0.00005	−0.00046	−0.00011
	(−2.25)	(−0.21)	(−2.47)	(−0.18)	(−2.42)	(−0.58)
A	−0.20738	0.12154	−0.25435	0.12908	−0.25713	−0.22645
	(−1.95)	(0.68)	(−2.32)	(0.74)	(−2.38)	(−1.71)
N	−0.01001	−0.27341	0.03228	−0.27116	0.03506	0.03029
	(−0.12)	(−1.92)	(0.36)	(−1.93)	(0.41)	(0.28)
N_U			0.13505		0.11312	0.83705
			(1.26)		(1.42)	(8.48)
$\begin{bmatrix}\sigma_I & \\ \rho_{UI} & \sigma_U\end{bmatrix}$	$\begin{bmatrix}1.5956 & \\ (11.5) & \\ 0.0453 & 2.6238 \\ (0.43) & (13.0)\end{bmatrix}$		$\begin{bmatrix}1.5664 & \\ (11.3) & \\ -0.0782 & 2.6101 \\ (-0.62) & (13.0)\end{bmatrix}$		$\begin{bmatrix}1.5663 & \\ (11.5) & \\ -0.0576 & 1.8769 \\ (-0.56) & (13.0)\end{bmatrix}$	
	Log $L = -409.3745$		Log $L = -408.6295$		Log $L = -378.5628$	

[a]Asymptotic *t*-statistics are displayed in brackets. They are computed from the analytic second derivatives.

namely, the value of the adjacent tract and that value squared, are the same in the informed and maximum uninformed bid equations. They are significant only in the bid equation of the informed firm. This is consistent with the prediction of the theoretical model that the bids of the non-neighbor firms are much "noisier" than the bids of the neighbor firms. Finally, note that the estimated standard error of the residuals in the informed bid equation is much lower than that of the maximum uninformed bid, although the informed bid itself has a higher standard deviation. (See Table 2.) We can explain a much higher percentage of the variation in the informed firm bids.

IV. A Competitive Bidding Model

In this section, we consider an alternative to the coordination model of neighbor firm bidding. We shall estimate a bidding model under the assumption that neighbor firms act independently and competitively. Our objective is to determine whether estimation under this behavioral hypothesis leads to implications which are not consistent with the theory of competitive bidding.

In the competitive bidding model, each neighbor firm observes a private signal on the value of the drainage tract, which, conditional on the value of the tract, is independently distributed across firms. The precisions of the signals are assumed to be identical, that is, information is symmetrically distributed among the neighbor firms. We shall continue to assume that neighbor firms view the bids of non-neighbor firms as uninformative random variables, and care only about the distribution of the maximum uninformed bid.[4]

---

[4]This assumption is somewhat *ad hoc*. In his analysis of an auction with asymmetric information in which uninformed firms have noisy, but private estimates of the informed firm's valuation, Wilson (1975) shows that firms which have access only to public information should never bid. If they do, they will lose money.

The equilibrium outcome depends on the nature of the neighbors' information. If the neighbors each know the exact value of the tract, then competition should lead to complete rent dissipation on multiple neighbor tracts. Table 4 indicates that the profit data are inconsistent with this sort of competition. Therefore, a competitive model can be consistent with the data only if we assume that neighbors' private information is imperfect.

The theoretical prediction concerning the sign of the equilibrium response of neighbors' bids to an increase in the number of neighbor firms is ambiguous: more competition causes the probability of winning to fall, which has a positive effect on bids. But, it also causes the expected value of the tract conditional on winning to fall, since there are now more observations on the value of the drainage tract. Therefore, the "winner's curse" is more acute, which has a negative effect on bids. Participation probabilities should decrease, however, since the reservation signal above which a neighbor firm participates is strictly increasing in the number of neighbor firms.

The ambiguity can be resolved for specific distribution functions of the private information. Smiley (1979) has shown that, for several distributions of the exponential class (including the lognormal), the equilibrium bid function is initially increasing in the number of bidders and then decreasing. The value at which the derivative changes sign depends upon the precision of the signal. For the degree of uncertainty present in our sample, the critical value is generally not less than four. Since there were only ten tracts that had four or more neighbor firms, this implies that, given plausible assumptions on the distribution functions, the neighbor firm's bid should be increasing in the range of the data.

The likelihood function for this model is defined as follows. Let I_t denote the index set for the neighbor firms on tract t. Define indicator variables for the informed firms,

$$Y_{it} = W_{it}'\theta_I + \varepsilon_{it}, \qquad i \in I_t,\ t = 1,\dots,T,$$

where W_{it} is a vector of regressors for tract t, θ_I is a parameter vector for informed firms, and the ε_{it}'s are jointly normal and independent (across tracts *and* firms) unobservable random variables. The indicator variable for the maximum uninformed bidder is defined as

$$Y_{Ut} = W_{Ut}'\theta_U + \varepsilon_{Ut}, \qquad t = 1,\dots,T,$$

where W_{Ut} is a vector of regressors for tract t, θ_{Ut} is a parameter vector for the maximum uninformed bid, and the ε_{it}'s are joint normal and independent (across tracts *and* firms) unobservable random variables. Bids are determined as follows: for each $i \in I_t$, and $i = U$,

$$\begin{aligned} \log(B_{it}/R_t) &= Y_{it} && \text{if } Y_{it} \geq 0 \\ &= 0 && \text{otherwise.} \end{aligned}$$

A tract enters the sample only if at least one bid is positive. Therefore, the likelihood function must account for both truncated random variables and sample selection issues. The latter is particularly important in the competitive model, since the likelihood of a tract entering our sample is higher when there are more neighbor firms, *ceteris paribus*.

Define dummy variables $d_{it} = 1$ if neighbor firm i bids on tract t, and $d_{it} = 0$ otherwise. Similarly, $d_{Ut} = 1$ if at least one nonneighbor firm bids on tract t, and $d_{Ut} = 0$ otherwise. Then the log likelihood function for the sample is given by

$$\begin{aligned} \log L = \sum_t \Biggl\{ & \sum_{i \in I_t} [d_{it} \log(\phi(\varepsilon_{it}/\sigma_I)/\sigma_I) \\ & + (1 - d_{it}) \log(\Phi(-W_{it}\theta_I/\sigma_I))] \\ & + d_{Ut} \log(\phi(\varepsilon_{Ut}/\sigma_U)/\sigma_U) \\ & + (1 - d_{Ut}) \log(\Phi(-W_{Ut}\theta_U/\sigma_U))] \\ & - \log\biggl(1 - \prod_{i \in I_t} \Phi(-W_{it}\theta_I/\sigma_I) \\ & \cdot \Phi(-W_{Ut}\theta_U/\sigma_U)\biggr)\Biggr\}. \end{aligned}$$

TABLE 9—JOINT DISTRIBUTION OF NEIGHBOR BIDS AND MAXIMUM NON-NEIGHBOR BID[a]

	Equation (1)		Equation (2)	
Independent Variable	Dependent Variable $\log(B_I/R)$	Dependent Variable $\log(B_U/R)$	Dependent Variable $\log(B_I/R)$	Dependent Variable $\log(B_U/R)$
Constant	1.7493	2.1856	1.7380	2.1165
	(2.516)	(2.292)	(2.49)	(2.22)
π	0.0971		0.0994	0.0363
	(3.40)		(3.47)	(0.97)
π^2	−0.00038		−0.00040	−0.00030
	(−1.19)		(−1.24)	(−0.68)
V	0.0400	−0.0015	0.0388	−0.0102
	(1.59)	(−0.04)	(1.54)	(−0.33)
V^2	−0.00049	−0.00004	−0.00046	0.00005
	(−1.46)	(−0.11)	(−1.40)	(0.15)
A	−0.2723	0.0626	−0.2759	0.3704
	(−1.94)	(0.33)	(−1.96)	(0.19)
NF	−0.7576	−0.4346	−0.7552	−0.4152
	(−3.96)	(−1.52)	(−3.93)	(−1.44)
$\begin{bmatrix} \sigma_I & \\ \rho_{UI} & \sigma_U \end{bmatrix}$	$\begin{bmatrix} 2.6747 & \\ (12.9) & \\ 0 & 2.7608 \\ & (10.6) \end{bmatrix}$		$\begin{bmatrix} 2.6793 & \\ (12.9) & \\ 0 & 2.7519 \\ & (10.6) \end{bmatrix}$	
	$\text{Log } L = -554.468$		$\text{Log } L = -553.781$	

[a] t-statistics are displayed in brackets. They are computed from analytic second derivatives.

The maximum likelihood estimates for the bid functions of the firms are reported in Table 9. The first two columns give the estimates of the joint distribution of bids when the conditioning variables for the maximum non-neighbor bid consist of the public information variables, and the neighbor bids are assumed to be a function of these variables, and our proxies for the neighbor firm's private information, tract profitability, and that variable squared. The final two columns give the estimates of the joint distribution when the tract profitability variables are included in the set of regressors for the maximum non-neighbor bid.

There are two ways in which the results reported in Table 9 are *not* consistent with the theoretical predictions of the competitive model. First, the bid function of the neighbor firm is strictly decreasing, not increasing, in the number of neighbor firms. This result is consistent with phantom bids, which are necessarily lower than the collusive bid, being often submitted on multiple neighbor tracts. In contrast, the *maximum* neighbor bid is independent of the number of neighbor firms. Second, the absolute value of the coefficients for the number of neighbor firms in the bid equations of the maximum uninformed bid are smaller than the corresponding coefficients in the bid equations for the neighbor firms. The opposite should be true, since an increase in the number of neighbor firms causes a larger decrease in the expected value of the tract conditional on the event of winning for the non-neighbor firm than the neighbor firm (i.e., the "winner's curse" is more acute for the less informed firms).

The results reported in this section may be consistent with a competitive bidding model in which one neighbor has access to superior information, and the others' information is essentially equivalent to that of the non-neighbors.

V. Conclusion

The data indicate that firms owning neighbor tracts have an informational advantage over non-neighbors in offshore drainage lease auctions. They exploit this advantage by shading their bids substantially below their expectation of the value of the tract. This translates into significantly higher returns, expressed as a percentage of discounted social value, than on wildcat tracts, where the distribution of information is relatively symmetric. The non-neighbors also account for their disadvantage, by bidding conservatively. As a consequence, they do *not* suffer from the winner's curse, but rather break even on average.

The pattern of neighbor bids appears to be inconsistent with a model in which they behaved competitively. Rather, the evidence indicates that they may have coordinated their bids to maximize joint returns. Such behavior may have been encouraged by the legality of joint bids, and by the presence of unitization agreements which could facilitate transfer payments.

The empirical results are also not consistent with a model in which the neighbor and non-neighbor firms are known to differ only in their extraction costs (i.e., roughly speaking, the distributions of neighbor and non-neighbor valuations differ in the first rather than the second moments). In that event, non-neighbors would have to bid cautiously, and as a result, would win fewer tracts. But their bids, conditional on participating, should be correlated with tract values. The evidence refutes this view, for only neighbor participation rates and bids, and *not* non-neighbor bids or their number, exhibit such a correlation.

A final alternative might be that firms adopt simple rule-of-thumb strategies. However, despite the complexity of calculating the Nash equilibrium strategy, the strategies themselves are relatively simple. Consequently, it is not implausible that well-financed bidders with experience in these auctions would employ optimal strategies.

APPENDIX

Proof of Theorem:

One can verify easily that the strategies specified above form an equilibrium. The proof of necessity is essentially the same as the one given by Engelbrecht-Wiggans, Milgrom, and Weber (EMW). Using their arguments, one can show that, in equilibrium, σ is nondecreasing, and that the expected payoff to the uninformed firm is zero. Furthermore, the bid distributions are atomless, except possibly at R and 0. Therefore, given any b contained in the support of G, $E[H-c-b|H<\tau(b); z]=0$. Substituting $b=\sigma(h)$, this implies

$$\text{(A1)}\quad \sigma(h)=E[H|H<h;z]-c$$

$$=h-c-\int_{-\infty}^{h}F(s;z)\,ds/F(h;z).$$

It then follows immediately from (A1) that the support of G (and the range of σ) is at most $\{0\}$ and $[R,\overline{H}-c]$.

Define

$$\hat{h}=\sup\{h|\sigma(h)=R\}$$

to be the largest valuation at which the informed firm bids R. Let λ denote the Lagrange multiplier associated with the constraint $b\geq R$. Optimality of $\sigma(h)$ implies that, given any h,

$$\text{(A2)}\qquad -G(b)+(h-b)G'(b)+\lambda=0,$$

$$\lambda\geq 0,\quad b\geq R.$$

Transforming variables and integrating the differential equation on (h,∞) for $h>\hat{h}$ yields

$$\text{(A3)}\quad \lim_{s\to\infty}\log G(\sigma(s))-\log G(\sigma(h))$$

$$=\int_{h}^{\infty}\frac{f(s;z)\int_{h}^{s}F(u;z)\,du}{cF(s;z)^{2}+F(s;z)\int_{h}^{s}F(u;z)\,du}\,ds$$

Applying the boundary condition $\lim_{s\to\infty}G(\sigma(s))=1$ implies

$$\text{(A4)}\quad G(\sigma(h))$$

$$=\exp\left\{-\int_{h}^{\infty}\frac{f(s;z)\int_{h}^{s}F(u;z)\,du}{cF(s;z)^{2}+F(s;z)\int_{h}^{s}F(u;z)\,du}\,ds\right\},$$

which is strictly positive. Since any bid less than R is rejected by the seller, we have that $G(0)=G(R)$. Thus, the event in which no non-neighbor firm bids occurs with positive probability.

Given any $h\in[R,\hat{h})$, λ is strictly positive, and the informed firm's optimal bid is R. At any $h<R$, it does not submit any bid. Therefore, the equilibrium bid distribution of the neighbor firm contains atoms at 0 and R, and the value of these atoms are $F(R)$ and $F(\hat{h})-F(R)$, respectively.

REFERENCES

Engelbrecht-Wiggans, Richard, Milgrom, Paul R. and Weber, Robert J., "Competitive Bidding and Proprietary Information," *Journal of*

Mathematical Economics, April 1983, *11*, 161–69.

Hendricks, Kenneth, Porter, Robert H. and Boudreau, Bryan, "Information, Returns and Bidding Behavior in OCS Auctions: 1954–1969," *Journal of Industrial Economics*, June 1987, *35*, 517–42.

Johnson, Norman L. and Kotz, Samuel, *Distributions in Statistics: Continuous Multivariate Distributions*, New York: Wiley & Sons, 1972.

Libecap, Gary D. and Wiggans, Steven N., "The Influence of Private Contractural Failure on Regulation: The Case of Oil Field Unitization," *Journal of Political Economy*, August 1985, *93*, 690–714.

Mead, Walter J., Moseidjord, Asbjorn and Sorenson, Philip E., "Competitive Bidding Under Asymmetrical Information: Behavior and Performance in Gulf of Mexico Drainage Lease Sales, 1959–1969," *Review of Economics and Statistics*, August 1984, *66*, 505–508.

Milgrom, Paul R. and Weber, Robert J., "The Value of Information in a Sealed Bid Auction," *Journal of Mathematical Economics*, June 1982, *10*, 105–14.

______, "Distributional Strategies for Games with Incomplete Information," *Mathematics of Operations Research*, November 1985, *10*, 619–32.

Smiley, Albert K., *Competitive Bidding Under Uncertainty: The Case of Offshore Oil*, Cambridge: Ballinger, 1979.

Weverbergh, M., "Competitive Bidding with Asymmetric Information Reanalyzed," *Management Science*, March 1979, *25*, 291–94.

Wilson, Robert J., "Competitive Bidding with Asymmetric Information," *Management Science*, July 1967, *13*, 816–20.

______, "On the Incentive for Information Acquisition in Competitive Bidding with Asymmetrical Information," Stanford University, mimeo., 1975.

______, "A Bidding Model of Perfect Competition," *Review of Economic Studies*, October 1977, *44*, 511–18.

Some Evidence on the Winner's Curse

By STUART E. THIEL*

The mathematical complexity of auction theory has restricted empirical work to qualitative tests of its basic predictions, with mixed results. This paper exploits the theory of order statistics to derive a Nash bid function that is linear in its parameters, allowing direct estimation by linear least squares. The model fits the data reasonably well, and does not provide evidence that the winner's curse is a significant problem in the highway construction industry.

In common-value auction bidders are competing to purchase an item whose value will only subsequently be revealed, as in offshore oil tracts or the rights to publish a book in paperback. The phenomenon known as the "winners's curse" turns up frequently in discussions of common-value auctions. (See John H. Kagel and Dan Levin, 1986; and surveys by R. Preston McAfee and John McMillan, 1987; and Richard Thaler, 1988.) While on close inspection, the concept can be quite confusing, the basic idea is this. In a common-value auction, each bidder must independently obtain an unbiased estimate of the unknown value of the object at auction. Then, even though the expected value of each estimate is correct, if each bidder bids his estimate the winner is nearly certain to have overestimated the true value, and so to bid too high. This is a result of the fact that even if the estimation procedure is unbiased, the bidder with the highest actual estimate (due to sampling error) is the winner. The winner's curse is an excellent example of decision making that fails to incorporate the actual statistical and economic process by which information is received by economic agents.[1]

Clearly the winner's curse is not consistent with our received notions of rationality and economic equilibrium. Indeed, in his seminal article on common-value auctions, Robert Wilson (1977) showed that optimizing behavior requires that bidders compensate for the potential bias by paying explicit heed to one another's behavior. In a Nash equilibrium the winner, far from being cursed, expects positive profits. If it is costless to submit a bid, the true value of the object should exceed the winning bid.

Wilson's result has not been borne out by the data. Studies of offshore oil leases indicate that the winner's curse does exist (E. C. Capen, R. V. Clapp, and W. M. Campbell, 1971; John Lohrenz, 1974; Kenneth Hendricks, Robert Porter, and Bryan Boudreau, 1987). So does the experimental evidence of Max Bazerman and William Samuelson (1983), although experiments by Kagel and Levin (1986) do suggest that experience can teach bidders how to avoid it.

If it exists, the winner's curse results from a failure of bidders to optimize correctly. In any discussion of such failure, it is important to distinguish between the bidder's estimation problem and his bidding problem. The distinction is not common in bidding theory, but it is advanced, forcibly, by Robert Stark and Thomas Varley (1983). A bidder might overbid because he overestimated the value, or because he overestimated his opponents' bids.

In this paper, I focus on the former interpretation. The winner's curse is avoided by any bidder using a valuation function whose expectation, *conditional on winning*, is an unbiased estimate of the object for sale. The function derived below requires that the bidders be individually rational in the sense of statistical decision theory, not attuned to one another as in game theory. The only belief each bidder must have about the others' actions is that the winner will be the firm with the largest estimate. The approach is to notice that mathematically, the bidding function maps the distribution of estimates into a distribution of bids, and the winner is the bidder with the first-order statistic from that sample. Knowing this, a bidder can offset the winner's curse by computing the expected difference between the first-order statistic and the value of the object for sale, and submit a bid that is shaved accordingly. This strategy guarantees that the bidder will earn zero-expected profits in equilibrium no matter how clever his opponents are.

---

*Department of Economics, Washington State University, Pullman, WA 99164–4860. The comments of Keith Brown, Ben Craig, Greg Duncan, Don Hausch, and those of the referees have led to great improvements in this work. Errors are mine.

[1] The definition of winner's curse has been controversial and the source of some debate. See James Cox and R. Mark Isaac, 1984, 1986, and Keith Brown, 1985, 1986.

But in a Nash equilibrium the winner will expect positive profits, and there must be an additional bit of strategy by which a bidder takes account of his opponents' behavior. Below, I exhibit a Nash bid strategy in which Wilson's bid-shaving is expressed as a separate additive term common to all bidders.

Using the theory of order statistics leads to a functional form that is linear in the parameters, inviting estimation by conventional econometric methods. In addition to ease of estimation, the method makes efficient use of the data by using all the bids, not only those of the winners. Further, the test of the winner's curse does not depend on (possibly flawed) estimates of the true value of the object at auction, but rather on properties of the bid strategy that are independent of that value.

The parameters are estimated using data from auctions for highway construction contracts. Generally, I find strong evidence that this model captures some elements of the firms' bid strategies, and that if the winner's curse exists, it may not be large.

Several theoretical papers have discussed the similarity of order statistics to auctions, but the only empirical use of which I am aware is by Lance Brannman, J. Douglass Klein, and Leonard Weiss (1987). They compare their regression results to tabulated values of expected order statistics, but do not use them directly in their estimation.

A difficulty with empirical work on the winner's curse is that the winner is only cursed relative to the true value of the item at auction, which is every bit as difficult for the econometrician to estimate as it was for the bidders. For offshore oil tracts, even after the test holes are drilled and the size of the pool is known, the present value of that resource depends in a complicated and uncertain way on the future price paths of oil and gas, confounding any direct claim that the winner did in fact over- or underbid.

For public works, such as my highway construction data, there is not even a market value against which to compare. Instead, one must use the agency's own estimate. Not only is that estimate just as uncertain as any bidder's, since all parties are professionals of presumably similar competence, but there is no particular reason to expect the agency's estimate to be unbiased. After all, if they go to the trouble to have a bid letting for a project, they must want to pave its way, tempting them to adjust their estimate.

The theory developed below attempts to avoid that problem by measuring the winner's curse in terms of two parameters that are explicitly independent of the true cost of the project. Unfortunately, technical econometric problems undermine the crispness of the test and lead to ambiguous results. I believe that the winner's curse is at most a slight problem in this industry, but the data admit other interpretations.

A further issue is that of censoring. It is intuitively plausible that there is an unobservable threshold value, such that a firm whose value estimate exceeds the threshold will not bid. Otis Gilley and Gordon Karels (1981), and Stanley Johnson and John Kuhlman (1983) have found evidence of such censoring in oil-lease and highway-contract bidding data, respectively. In the present paper the evidence is mixed. Imposing the hypothesis of censoring yields coefficient estimates giving near-perfect corroboration of the theory, but there is reason to suspect that that version of the model is misspecified.

A word must be said about corrupt practices like bid-rigging, for which the highway construction industry is notorious. Bid-rigging is difficult to detect: that is the whole idea. The conspirators have every incentive to devise bids that mimic honest bids. Kuhlman and Johnson (1984) suggest that, once bid-rigging has become suspected, elaborate comparisons of suspect contracts to similar contracts can be used as legal evidence. Jonathan Feinstein, Michael Block, and Frederick Nold (1985) work directly with a panel of data from one state, aided by the *ex post* information of which firms in the panel were later convicted of bid-rigging.

My data, a cross section from 33 states, do not lend themselves to either method. I do find that in the states in which the agency's value estimate is announced in advance, there is a wholesale breakdown of the model, which might indicate collusion, but is more likely evidence that the theory is inappropriate to such cases. Also, in the other states there is a faint indication that the winner might have access to some extra information not available to the other bidders or to the researcher. Generally, however, all I can say about bid-rigging is that if it is there, it is well camouflaged.

I. The Bidding Model

A. *Order Statistics*

First I present a few facts about order statistics necessary for the discussion below. Details can be found in the text by H. A. David (1981), especially in Chapters 1 and 6.

Consider a random sample $\{X_j\}$ of size N drawn from a continuous distribution. For the present purpose, assume the parent distribution $F(x|\theta,\sigma)$, with density $f(x|\theta,\sigma)$ is in the location-scale family. The location parameter θ and the scale parameter $\sigma > 0$ may be, but are not necessarily, the mean and standard deviation. Then X admits the standardization $Z = (X-\theta)/\sigma$, where the distribution $G(z)$, with density $g(z)$, is independent of θ and σ. Of use below will be the facts $F(\theta|\theta,\sigma) = G(0)$ and $f(\theta|\theta,\sigma) = g(0)/\sigma$; the latter is obtained by simple differentiation.

Arrange the sample in increasing order, and let the subscript $j{:}N$ denote the jth order statistic in a sample of size N: $X_{1:N} < X_{2:N} < \cdots < X_{N:N}$. Clearly $X_{j:N} = \theta + \sigma Z_{j:N}, E(X_{j:N}) = \theta + \sigma E(Z_{j:N})$, and $\mathrm{Cov}(X_{i:N}, X_{j:N}) = \sigma^2 \mathrm{Cov}(Z_{i:N}, Z_{j:N})$. Note that although the original data formed a random sample, arranging them in order changed the joint distribution.

Write $E(Z_{j:N}) = \zeta_{j:N}$; in general,

$$(1) \quad \zeta_{j:N} = N\binom{N-1}{j-1} \times \int sG(s)^{j-1}[1-G(s)]^{N-j}dG(s).$$

The expression for $\mathrm{Cov}(Z_{i:N}, Z_{j:N}) = r_{ij:N} > 0$ is even more complicated. Fortunately, since $G(Z)$ is independent of θ and σ, for any given distribution the values of $\zeta_{j:N}$ and $r_{ij:N}$ can be tabulated. If $G(Z)$ is the normal distribution, as will be needed below, William Beyer (1966) has tabulated $\zeta_{j:N}$ for $2 \le N \le 39$ and $r_{ij:N}$ for $2 \le N \le 20$.

Suppose we need an unbiased estimate of σ. Inspection of $E(X_{j:N}) = \theta + \sigma\zeta_{j:N}$ suggests regressing $X_{j:N}$ on $\zeta_{j:N}$ and an intercept. Ordinary least squares (OLS) will give estimates[2] $\hat{\theta}$ and $\hat{\sigma}$ that are unbiased, but not efficient because $\mathrm{Cov}(X_{i:N}, X_{j:N}) = \sigma^2 r_{ij:N}$. However, the covariance matrix $\sigma^2 R_N = \sigma^2[r_{ij:N}]$ is known up to σ^2, and we can use that information to perform generalized least squares (GLS) using the exact Aitken transformation.

That method is at the heart of this paper. Three further facts that will come up again are: David cites A. K. Gupta (1952) to the effect that if the parent distribution is normal, the efficiency loss from using OLS instead of GLS is slight; the standard error of the regression gives a second, but biased, estimate of σ; and since the regression errors have come by way of the order statistics, they do not share the parent probability distribution. In particular, even if $F(x|\theta,\sigma)$ is the normal distribution, the errors are not themselves normally distributed.

B. *A Linear Bid Stategy*

Consider a sealed-bid auction and a group of N bidders, each of whom is risk-neutral and wishes to maximize expected profits.

---

[2] In general, the circumflex (ˆ) notation refers to the estimate, for example, $\hat{\sigma}$ is the estimate of σ.

Each has an independent estimate, X_j, $j = 1,2,\ldots, N$, of θ, the unknown value of the object at auction, and the high bid wins. The values $\{X_j\}$ are drawn from a common distribution $F(x|\theta, \sigma)$; all bidders know N, the functional form of $F(x)$, and $\sigma > 0$. $F(x)$ is continuous and a member of the location-scale family.[3]

Each bidder regards X_j as his best point estimate of θ. That is, either the bidders are all classical statisticians, forced by their methodology to heed only the samples they have taken, or they are all Bayesians, sharing the same posterior distribution. In either case, each bidder believes $\hat{F}(X_j|X_j, \sigma) = F(\theta|\theta, \sigma) = G(0)$ and $\hat{f}(X_j|X_j, \sigma) = f(\theta|\theta, \sigma) = g(0)/\sigma$. This is a crucial step in the derivation, because it is this insight that causes the bid strategy to be linear.

Suppose an individual bidder's first priority is to avoid the winner's curse. He wants to find a valuation function $p_N(x)$ with the property that if he wins, he breaks even: $E(p_N(X)|\text{win}) = \theta$. If the equilibrium bid function is monotone increasing, the bidder with the highest estimate is the winner. His point estimate must have been $(X_{N:N}) = \theta + Z_{N:N}$, with expectation $E(X_{N:N}) = \theta + \sigma\theta_{N:N}$. Thus if $p_N(X) = X - \sigma\zeta_{N:N}$,

$$E(p_N(X)|\text{win}) = E(X_{N:N}) - \sigma\zeta_{N:N} = \theta + \sigma\zeta_{N:N} - \sigma\zeta_{N:N} = \theta,$$

as required.

Here, we concentrate on symmetric equilibria, in which all bidders play the same strategy. Wilson (1977) showed that in a symmetric Nash equilibrium the winner expects positive profits. Let $b(X)$ be an individual's bid, and let B be the bidder's highest competing bid, with subjective prior distribution $H_N(B)$ and density $h_N(B)$. The well-known first-order condition for a Nash equilibrium (see, for example, Gilley and Karels, 1981) is

$$(2) \qquad b(X) = p_N(X) - \frac{H_N(b(X))}{h_N(b(X))},$$

where $E(p_N(X)|\text{win}) = \theta$. From that we get the following proposition.

PROPOSITION 1: *Let* $p_N(X) = X - \sigma\zeta_{N:N}$ *and* $W(N) = [G(0)/g(0)]/(N-1)$. *Then*

$$(3) \qquad b(X) = p_N(X) - \sigma W(N)$$

gives a symmetric Nash equilibrium in pure strategies.

PROOF:

To prove the proposition, we must show that $b(X)$ simultaneously satisfies everyone's first-order condition. First, note that $b'(X) = 1 > 0$, so $b(X)$ is monotone increasing, and that $E(p_N(X)|\text{win}) = \theta$ as above.

Second, $H_N(b(X))$ is the probability that $b(X)$ is greater than the highest competing bid. But, again because $b(X)$ is order-preserving, it must equal the probability that X is the highest signal, so $H_N(b(X)) = F(X)^{N-1}$. Differentiating both sides and using $b'(X) = 1$ gives $dH_N(b(X))/dX = h_n(b)b'(X) = h_N(b) = (N-1)f(X)^{N-2}$ so

$$H_N(b)/h_n(b) = [\hat{F}(X)/\hat{f}(X)]/(N-1).$$

Using $\hat{F}(X|X\sigma) + G(0)$ and $\hat{f}(X|X, \sigma) = g(0)/\sigma$ gives $\sigma W(N)$.

Thus the bid function satisfies the first-order conditions for all bidders at once. Since $b'(X)$ is finite, the resulting Nash equilibrium is in pure strategies.

The expected value of the winning bid is $\theta - \sigma W(N)$, so $\sigma W(N) > 0$ is the expected profit. $W(N)$ converges monotonically to zero as N gets large, so the winning bid converges to the true value, at rate $1/N$, as derived by Wilson.

Thaler points out that in experiments the average bid is typically significantly less than θ. He echoes the literature in associating that observation with risk aversion and the winner's curse. However, neither of

---

[3] A referee has pointed out that the entire method of the paper can be adapted to distributions that can be transformed into location-scale distributions, for example, the lognormal. The lognormal is especially important in the study of offshore oil lease bidding; see James Ramsey (1983) and Robert Wilson (1983).

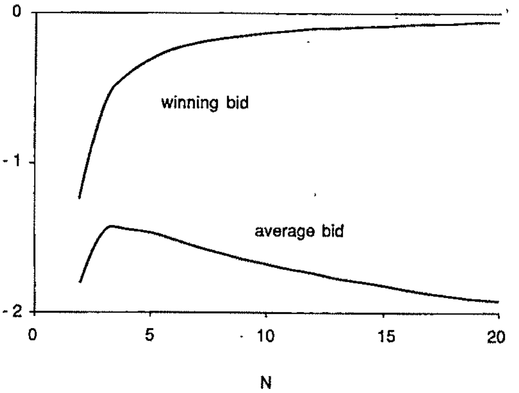


FIGURE 1

those explanations is necessary: from (3), $E(b(X)) = \theta - \sigma\zeta_{N:N} - \sigma W(N) < \theta$. Also, since $\zeta_{N:N}$ is increasing in N, as N gets large eventually the $W(N)$ term becomes dominated by $\zeta_{N:N}$ and $E(b(X))$ diverges from θ. Figure 1 shows the expected winning and average bids when $F(x|\theta, \sigma)$ is the standard normal distribution.

Note also that this bid strategy consists of simply adding a constant to the value estimate X, so the bid and signal distributions differ only in location (compare Bazerman and Samuelson, 1983). That constant is a function only of N and σ and can easily be arranged in a table or programmed into a calculator for quick consultation by the bidders. As Figure 1 and (3) indicate, the constant has two parts. The first part is a statistical correction for the winner's curse that increases with the number of bidders. The second is a part that decreases with the number of bidders and reflects the economic effects of increased competition.

The empirical work of this paper concerns auctions for public works contracts, in which the low bid wins. By symmetry, the winner has the lowest signal, $X_{1:N}$, and $\zeta_{1:N} < 0$. The appropriate bid strategy is

$$(4) \qquad b(X) = X - \sigma\zeta_{1:N} + \sigma W(N).$$

Since the bidders cannot know their ranking at the time that they bid, the jth bid has expectation

$$(5) \qquad E(b_j) = \theta + \sigma\zeta_{j:N} - \sigma\zeta_{1:N} + \sigma W(N).$$

If a suitable unbiased estimator of θ can be found, this form suggests estimation of the bid parameters by linear least squares. Here is where it was necessary to assume $F(x|\theta, \sigma)$ in the location-scale family; since the values of $\zeta_{j:N}$ and $W(N)$ are independent of θ and σ, they can be found in tables or calculated for use as regressors.

Suppose that not all samplers bid. There are N^* samplers, of whom $N \leq N^*$ actually bid, and both N^* and N are known to the bidders at the time of the auction. Clearly $p_{N^*}(X)$ will depend on the sample size, but the term $W(N)$, explicitly accounting for the number of rivals, continues to depend only on N. If N^* cannot be observed, the sample is censored, posing difficulties for the estimation. That possibility is considered below.

II. The Data

A. *Description*

Highway construction contracts are let to the qualified firm entering the lowest bid. The states and bonding companies impose prequalification conditions on the bidders, ensuring that the firm is competent to do the job and concentrating small firms on small jobs, large firms on large jobs. Thus the firms in each auction will tend to be homogeneous.

On the day of the bid letting, the firms' representatives commonly congregate at the agency's office and compare notes on their estimates of the value of the contracts. Each bidder can learn the number of rivals he has for a particular contract, and the process generates a pool of public information about its value, to which the bidder adds the information he has kept private to form a value estimate. These value estimates, then, can be regarded as independent observations from an appropriately defined Bayesian posterior distribution for θ.

The highway agency also has estimated the value of the contract. Most states keep their engineer's estimate secret before the bid letting. The few that do not cause drastic amendments in the bid strategies. I do not offer a theory of those auctions, but do show that data from such states do not fit the model I develop.

The data are a cross section of 150 auctions from 33 states; 20 of the auctions are from five states[4] that announce the engineer's estimate in advance. The data were obtained from the weekly feature "Unit Prices" in the construction industry trade magazine *Engineering News-Record* for the years 1982–85; unit prices are the prices of the individual items used in construction work. Let the auctions be indexed by $k, k=1,2,\ldots,T$. The variables in the data set are: the state; N_k, the number of bidders; m_k, the number of different items used in the project; b_{jk}, $j=1,2,H$, the low (winning), second, and high bids; and V_k, the state's engineer's estimate of the cost of the project.[5]

Since the bids from the five unusual states differ so strongly from the rest, most of the discussion below involves the 130 auctions untainted by the state's disclosure of its estimate. For those auctions the median value of a contract is $7.3 million, with a fairly even spread between small, medium, and large contracts. The number of bidders in an auction ranges between 3 and 16.

The jth bidder in the kth auction estimates the unit prices, multiplies by the amount of each item necessary to do the job, and sums to compute the estimate X_{jk}. Stark (1974) has noted that for a long-term project, considerations of optimal cash flow might make it desirable to overbid on some items and underbid on others, always keeping in mind that it is the sum that determines the winner. He provides a linear program for computing the optimal value estimate.

Whether the firm takes Stark's advice or not, the value estimate X_{jk} is the sum of many random variables. The number of summed variables is related to the number of different items, the amount of each item, and the cost of each item used in the project. Only 26 of the 150 contracts involved fewer than 30 items. If we assume that these variables are unbiased, uncorrelated, and have finite variance, the classical central limit theorem asserts that X_{jk} converges in distribution to a normal random variable whose mean is the value, θ_k, of the contract. I assume that the variance of X_{jk}, σ_k^2, is proportional to the number of items, m_k: $\sigma_k^2 = \phi^2 m_k$, and that the bidders know ϕ and observe m_k.

B. *The Engineer's Estimate*

The value of the contract, θ_k, cannot be observed and we must use V_k, the engineer's estimate, as a proxy. This raises the question of just what the engineer's objective is. That is, suppose that the state's engineer constructs his estimate from the same distribution as do the firms, so $V_k = \alpha_0 + \alpha_1 \theta_k + \omega_k$ and the error ω_k is independently normally distributed with mean zero and variance σ_k^2.

If the engineer's purpose is merely to provide an unbiased estimate of θ_k, then $\alpha_0 = 0$ and $\alpha_1 = 1$. However, in most states, V_k's statutory function is as a reservation price; officially, if no bidder meets the price, the project should go unbuilt. In practice, Johnson and Kuhlman (1983) report that if the lowest bid is not low enough, the state restudies its estimate in the light of the new information and often relents, letting the bid after revising its estimate. It seems likely that to avoid the extra effort, the engineer might want to overestimate θ_k, to help see to

[4] Illinois, Louisiana, Massachusetts, Nevada, Texas. There may be others; I made contact only with the states in the data set.

[5] Some auctions had to be culled from the sample: those for which $N \leq 2$ and $N > 20$, due to the limitations of the tables; those for which no engineers' estimate was included; and those for federal jobs, mostly U. S. Army Corps of Engineers harbor and river projects. The last group was eliminated because the projects were huge and dominated by far-flung consortiums, incompatibly with the state projects. Auctions for which the high bid was omitted were retained. A copy of the data set and a sample SAS program can be sent upon request either on paper or via electronic mail to any BITNET address.

it that the bidders meet the minimum and the contract is let the first time, suggesting $\alpha_0 > 0$. So long as $N > 1$, such padding does not help the bidders, whose main concern is still with each other, but it does complicate the estimation. The issue arises because, as will be seen below, using V_k in the estimation leads to results that are hard to interpret.

III. Econometric Specification

To proceed, under the hypothesis that the sampling distribution is normal, $G(0)/g(0) = \sqrt{\pi/2}$, so $W(N) = \sqrt{\pi/2}/(N-1)$. Indexing auctions by k, the three estimating equations for the kth auction have the form

$$(6)\quad b_{jk} = \theta_k + \beta_2\sqrt{m_k}\,\zeta_{jk} - \beta_3\sqrt{m_k}\,\zeta_{1k} + \beta_4\sqrt{m_k}\,W_k + \varepsilon_{jk},$$

$j = 1,2,H$, $k = 1,2,\dots,T$, where $\zeta_{jk} = \zeta_{j:Nk}$, $W_k = W(N_k)$ and ε_{jk} is a random error term. The covariance matrix of $(\varepsilon_{1k}, \varepsilon_{2k}, \varepsilon_{Hk})'$ is given by $\phi^2 m_k R_k$, where $R_k = [r_{ij:Nk}]$, $i, j = 1,2,H$, is the matrix of standardized covariance terms.

Written in the form of (6), the Nash-bidding hypothesis is that $\beta_2 = \beta_3 = \beta_4 = \phi$. The coefficient β_2 reflects a property of the sample, not of the theory, and shall be taken to be the most reliable estimate of ϕ. Let the winner's curse (WC) measure be $\mathrm{WC} = 1 - \beta_3/\beta_2$, so perfect hedging gives $\mathrm{WC} = 0$ and naive bidding gives $\beta_3 = 0, \mathrm{WC} = 1$. Firms have failed of adequate hedging against the winner's curse if $\mathrm{WC} > 0$. They have failed of proper Nash-bidding behavior if $\beta_4 \neq \beta_2$.

θ_k is unobservable. The relation discussed above, $V_k = \alpha_0 + \theta_k + \omega_k$, gives $\theta_k = \beta_0 + \beta_1 V_k - \omega_k, \beta_0 \leq 0$, $\beta_1 = 1$, as a proxy. This introduces the possibility of errors-in-variables bias, which will be discussed in Section IV. Substituting in to (6) gives triples of equations whose error vectors have covariance matrix $\sigma_k^2(R_k + I)$, where I is the 3×3 identity matrix. Weighting each equation by $1/\sqrt{m_k}$ to give the heteroskedasticity correction for $\sigma_k^2 = \phi^2 m_k$ results in transformed estimating equations

$$(7)\quad \frac{b_{jk}}{\sqrt{m_k}} = \frac{\beta_0}{\sqrt{m_k}} + \frac{\beta_1 V_k}{\sqrt{m_k}} + \beta_2\zeta_{jk} - \beta_3\zeta_{1k} + \beta_4 W_k + \eta_{jk},$$

where the covariance of the vector $(\eta_{1k}, \eta_{2k}, \eta_{Hk})'$ is $\phi^2(R_k + I)$. The weights on the coefficient estimates $\hat{\beta}_2, \hat{\beta}_3$, and $\hat{\beta}_4$ are canceled out because they, too, are estimates of the standard deviation of the original sampling distribution. Premultiplying both sides of the system (7) by $(R_k + I)^{-1/2}$ gives the exact Aitken transformation, after which each auction in the data set has covariance matrix $\phi^2 I$ and the system can be estimated efficiently by ordinary least squares.

Even though the joint distribution of the bids is no longer normal, the parameters are estimated by stacking the bid vectors from all of the independent auctions into one regression. Thus the coefficient estimates are linear functions of many such bid vectors and are approximately normal.

IV. Estimation

A. *Regression Results*

The regression results are reported in Table 1. Column 1 uses the entire data set of 150 auctions. The states were divided according to whether the engineer's estimate is announced in advance, by appropriate slope dummies on ζ_{jk}, ζ_{1k}, and W_k. Columns 2 and 3 result from partitioning the data set and running the regressions for the two kinds of state separately. The number of observations is less than triple the relevant number of auctions because in seven auctions the high bid was not reported in the data set.

In all cases, $\hat{\beta}_0 < 0$, although insignificantly in the states where the estimate is made known. In the other states, the indication is either that the winner is cursed to the tune of nearly \$1 million, on average, or that the engineers consistently overestimate by that amount. I have argued for the latter, but cannot insist dogmatically.

TABLE 1—GLS ESTIMATES OF UNCENSORED SPECIFICATION[a]

		Column 1	Column 2	Column 3
Regressor	Coefficient	All Bids	V_k Known	V_k Unknown
Constant	β_0	−0.8725	−0.2442	−0.9955
		(0.345)	(2.026)	(0.280)
V_k	β_1	0.9683	0.9718	0.9610
		(0.012)	(0.032)	(0.015)
ζ_{jk}	β_2	0.1267)	0.4411	0.1267
		(0.015)	(0.096)	(0.114)
ζ_{1k}	β_3	0.1046	−0.0046	0.1190
		(0.036)	(0.207)	(0.030)
W_k	β_4	0.2621	0.1241	0.2893
		(0.136)	(0.697)	(0.104)
$\zeta_{jk} D_k$		0.3188		
		(0.047)		
$\zeta_{1k} D_k$		−0.0547		
		(0.066)		
$W_k D_k$		−0.0211		
		(0.301)		
R^2		0.9713	0.9740	0.9679
Standard Error of Regression		0.3064	0.6511	0.2282
Number of Observations		443	54	389

[a] Standard errors are in parentheses. Dollar values are in millions. States are categorized by whether the state's estimate V_k is known in advance. In column 1, all states are used and the slope dummy D_k takes the value 1 if V_k is known; columns 2 and 3 result from partitioning the data set.

Also in all cases, $\hat{\beta}_1$ is significantly less than 1. That does not rule out the possibility that $\beta_1 = 1$, because as the coefficient estimate on a proxy variable, $\hat{\beta}_1$ is almost certainly biased downward. The standard error of $\hat{\beta}_1$ is quite small because the regression is nearly one of random variables on their mean, accounting for the high R^2 as well.

$\hat{\beta}_3$ and $\hat{\beta}_4$ are the strategy coefficients to be scrutinized. $\hat{\beta}_3$ indicates the degree to which the value estimate has compensated for the potential for the winner's curse, which is zero when $\beta_2 = \beta_3$. $\hat{\beta}_4$ measures the extent to which the firms are Nash bidders; ideally, $\hat{\beta}_4 = \hat{\beta}_2$.

One thing is simply obvious. Bidders in states that announce the engineer's estimate in advance do not use the strategy described above. Neither of the strategy coefficients $\hat{\beta}_3$ or $\hat{\beta}_4$ can be distinguished from zero. Since the theory was explicitly derived under the assumption that the seller had no reservation price, the result is not surprising. In this case the optimal strategy is probably a more complicated function than the one derived.

Paul Milgrom and Robert Weber (1982) have shown that the seller maximizes revenue by announcing its value estimate. The prediction of maximum revenue cannot be tested here, but they also note that the announced estimate might as well be the truth, since any systematic lie will be detected by the bidders. This implies that V_k is, in this case, likely to be an unbiased estimate of θ and the bidders, knowing V_k, have no reason to adjust for the winner's curse. That in turn suggests $\beta_0 = 0$, $\beta_1 = 1$, and $\beta_3 = 0$, all of which are compatible with the estimates in column 2. On the other hand, the coefficient on the profit term W_k, while not significant, supports the claim that even in these states the winner's expected profit is positive, of order $1/N$, and of similar magnitude to that earned in other states.

Auctions in the states keeping V_k a secret strongly corroborate the model. The winner's curse is small; in column 3 the measure WC $= 1 - \hat{\beta}_3/\hat{\beta}_2 = 6$ percent. As to $\hat{\beta}_4$, the Nash-bidding coefficient, first observe that its sign and standard error are evidence in favor of Wilson's original common-value model: anticipating the win, bidders allow themselves some extra profit. The suggestion is that they pad their bids more than double the level predicted. On the other hand, $\hat{\beta}_4$ compares well to the standard error of the regression, an independent estimate of ϕ.

Neither of the hypotheses $\beta_2 = \beta_3$ or $\beta_2 = \beta_3 = \beta_4$ can be rejected. However, formal hypothesis testing is not very powerful here, both methodologically, because the specification makes it natural to test the favorable outcome as the null hypothesis, and because the standard errors are too large to be helpful.

Using dummy variables to divide β_4 into the three coefficients β_{41}, β_{42}, and β_{4H}, the Nash coefficients for the low, second, and high bidders, gives estimates $\hat{\beta}_{41} = 0.353$, $\hat{\beta}_{42} = 0.271$, and $\hat{\beta}_{4H} = 0.243$ (standard errors all about 0.14). The winning bidder seems consistently able to bid slightly higher, relative to his estimate, than are the others.[6] While that might be evidence of illegal collusion, it more likely indicates some heterogeneity among firms or some other flaw in the strict common-value specification.

B. *Censoring*

Several authors, for example, Gilley and Karels (1981), and Johnson and Kuhlman (1983), have raised the issue of censoring, in that more firms make value estimates than bid. Strict classical censoring, in which there is some threshold such that only those with value estimates less than that figure choose to bid, is intuitively appealing and is the approach taken by those authors. It should be recognized, however, that that possibility is explicitly ruled out by the common-values model. The only way a firm can decide if its estimate is too high is to have some other estimate as a standard of comparison. But in the common-values model, the bidder is allowed to observe only his private signal X_j. If there is some public estimate, or if all bidders have access to similar extra private information, then the relevant probability distribution is the Bayesian posterior distribution obtained after the bidders individually devise new estimates by a suitable update rule.

Unfortunately, that posterior distribution cannot be observed by the researcher, so it is sensible to investigate censoring with respect to the sampling distribution. The order-statistical method of this paper is well-adapted to estimates from such censored distributions (Keith Brown and Deborah Brown, 1986). Suppose q percent of all samplers bid. Then if we have N bidders, we expect the original sample to have been of size $N^* = N/q$ (rounded to the nearest integer). The first and second bidders are now assigned expected order statistics based on N^*, $\zeta_{1:N^*}$, $\zeta_{2:N^*}$, and if there is classical censoring, the high bidder is merely the Nth of N^*, with $\zeta_{N:N^*}$. As discussed above, the number of actual bidders can be observed on bid-letting day, so $W(N)$ continues to be computed from N_k.

Johnson and Kuhlman (1983) estimate that in highway contracting somewhere between 35 and 50 percent of all samplers bid. In the estimates in Table 2, $N^* = N/0.41$, 41 percent being chosen due to the limitations of the tables of order statistics. Also, because the covariance terms are unavailable for $N^* > 20$, all estimates in Table 2 are ordinary least squares.

Column 1 gives, for comparability, the OLS estimates of the original specification. Note in passing that the standard error of the regression is slightly larger here than in the GLS estimates in column 3 of Table 1, corroborating Gupta's claim that the efficiency loss from using OLS is small, but that the parameter estimates here give much smaller measure of the winner's curse, WC $= 2$ percent.

Column 2 gives the censored estimates. The fit of the estimates in column 2 is almost too good; not only are the three coefficient estimates $\hat{\beta}_2, \hat{\beta}_3$, and $\hat{\beta}_4$ virtually equal, but

---

[6] Using dummies to make similar adjustments to $\hat{\beta}_2$, and $\hat{\beta}_3$ revealed no conspicuous differences in the coefficients.

TABLE 2—OLS ESTIMATES OF ALTERNATIVE SPECIFICATIONS

Regressor	Coefficient	Column 1 Original spec.	Column 2 Censored spec.	Column 3 Differenced spec.
Constant	β_0	-1.0239	-1.0309	
		(0.267)	(0.271)	
V_k	β_1	0.9595	0.9591	
		(0.014)	(0.014)	
ζ_{jk}	β_2	0.1265	0.2339	0.1289
		(0.012)	(0.022)	(0.013)
ζ_{1k}	β_3	0.1238	0.2270	0.1132
		(0.029)	(0.280)	(0.011)
W_k	β_4	0.2887	0.2729	
		(0.094)	(0.094)	
R^2		0.9704	0.9702	0.5405
Standard Error of Regression		0.2722	0.2730	1.972
Number of Observations		389	389	259

[a] Standard errors in parentheses. Dollar values in millions. Column 1 gives OLS estimates of the GLS estimates in column 3 of Table 1, included here for comparison. The estimates in column 2 are from the censored specification described in the text. Those in column 3 result from differencing the original system to eliminate V_k and W_k.

so is the standard error of the regression. On its face, that seems to be strong evidence of the joint hypothesis that the model is valid and there is censoring of the magnitude claimed by Johnson and Kuhlman. However, the caveat that the OLS estimate might underestimate the winner's curse carries down from the previous paragraph; also, in the next section it will be argued that the censored model is poorly specified.

C. *Diagnostics*

Two specification issues are the extent of errors-in-variables bias introduced by the proxy variable V_k, and the issue of whether the censored or uncensored specification is the correct one. Inspection of equation (1) shows that the regressors $\zeta_{j:N}$ are highly nonlinear functions of N. In general, as Jerry Hausman (1983) has explained, asymptotically consistent specification tests of nonlinear models require complicated iterative maximum likelihood techniques. The following method is offered as a simple alternative.

Since V_k is suspect, rearrange equation system (7) and eliminate V_k, obtaining the differenced system

$$\text{(8a)}\qquad \frac{E(b_{2k}-b_{1k})}{\sqrt{m_k}}=\beta_2\zeta_{2k}-\beta_3\zeta_{1k},$$

$$\text{(8b)}\qquad \frac{E(b_{Hk}-b_{1k})}{\sqrt{m_k}}=\beta_2\zeta_{Hk}-\beta_3\zeta_{1k}.$$

In the transformation, $\beta_4 W_k$ was also lost, but we can still investigate the winner's curse, because the term $\beta_3\zeta_{1k}$ in (8) measures the hedging strategy played by the bidders.

The OLS estimates of (8) are presented in column 3 of Table 2. In the original GLS specification, the winner's curse was 6 percent, and in the OLS specification, it is 2 percent. Here, eliminating V_k gives WC = 13 percent, suggesting that while there is some specification bias, the winner's curse is minimal in any case.

To address the censoring issue, note that from Table 2, in either the censored or uncensored specification, the winner's curse is quite small. Suppose WC = 0, implying, in equation (8), $\beta_2=\beta_3$. Then if the model is correctly specified, the ratio ϕ_a/ϕ_b of the parameters

$$\text{(9a)}\qquad \frac{E(b_{2k}-b_{1k})}{\sqrt{m_k}}=\phi_a(\zeta_{2k}-\zeta_{1k})$$

$$\text{(9b)}\qquad \frac{E(b_{Hk}-b_{1k})}{\sqrt{m_k}}=\phi_b(\zeta_{Hk}-\zeta_{1k})$$

should be equal to 1. If the ratio is different between the two specifications, that difference is due to the values of ζ_1, ζ_2, and ζ_H. There are two sets of ζ values, one for the original specification and one for the censored case.

Using the ζ values from the original model, the OLS estimates of (9a) and (9b) are $\hat{\phi}_a = 0.092$, $\hat{\phi}_b = 0.123$, with a ratio of 0.748. For the censored specification, the analogous estimates are $\hat{\phi}_a = 0.117$, $\hat{\phi}_b = 0.223$, with a ratio of 0.525. Thus for all its apparent corroboration of the theory, the censored model is evidently the worse of the two.

On inspection, the reason for the good fit of the censored model in Table 2 is obvious. One way to double, say, the estimates $\hat{\beta}_2$ and $\hat{\beta}_3$ in column 1 is to divide the regressors ζ_{jk} and ζ_{1k} by 2. But since $\zeta_{1:N^*}$, $\zeta_{2:N^*}$, and $\zeta_{N:N^*}$ are all similar monotone-decreasing functions of N^*, that is essentially what has been accomplished by substituting $N^* = N/0.41$. With better tables, the result could have been tuned even more finely. While it is interesting that the factor of 41 percent is close to that specified by Johnson and Kuhlman, the evidence is at best suggestive. This model can help to characterize bid strategies, but it should not be oversold.

V. Conclusion

Ambiguity about the exact specification should not obscure the positive results. This model accounts for regularities found in experimental work on common-value auctions. Bidders on highway contracts do seem to shave their bids in order to avoid the winner's curse, and the extent of the shaving does vary with the competition as predicted. In the field data, in those states where it would be surprising for the model to hold, because there is public information on project costs, it does not.

In those states whose highway auctions more closely match the hypotheses of the model, the coefficients on the bid variables have the right sign, they are significant, and we cannot reject the hypothesis that they are, as predicted, equal. Estimates of the winner's curse indicate that bidders' strategies differ from the optimal by from 2 to 13 percent, but these estimates are not precisely determined. The evidence here indicates that the winner's curse does not appear to be a significant problem in highway construction contracting, as it should not in economic equilibrium, but this is partly a matter of judging how large a discrepancy is "large." Bidders do seem to approximate the hypothesized Nash-bidding behavior. If anything, the contractors' profits are larger than predicted, not smaller.

The essential strength of this model is the same as its essential weakness: the estimating equations are specified, right down to functional form, by the theory. For the econometric specification to be so completely determined a priori is unusual in economics. For that specification actually to provide a rough fit to the data is even less common.

REFERENCES

Bazerman, Max and Samuelson, William, "The Winner's Curse: An Empirical Investigation," in R. Tietz, ed., *Aspiration Levels in Bargaining and Economic Decision Making*, New York: Springer-Verlag, 1983, 186–200.

Beyer, William H., ed., *CRC Handbook of Tables for Probability and Statistics*, 2nd ed., Cleveland: Chemical Rubber Co., 1966, 258–65.

Brannman, Lance, Klein, J. Douglass and Weiss, Leonard, "The Price Effects of Increased Competition in Auction Markets," *Review of Economics and Statistics*, February 1987, *69*, 24–32.

Brown, Keith C., "The Winner's Curse: What It Is and Why It Matters," Purdue University, Krannert School Discussion Paper No. 888, 1985.

______, "In Search of the Winner's Curse:

Comment," *Economic Inquiry*, July 1986, *24*, 513–16.

______ **and Brown, Deborah,** "Using Order Statistics to Estimate Real Estate Bid Distributions," *Management Science*, March 1986, *32*, 289–97.

Capen, E. C., Clapp, R. V. and Campbell, W. M., "Competitive Bidding in High-Risk Situations," *Journal of Petroluem Technology*, June 1971, *23*, 641–53.

Cox, James C. and Isaac, R. Mark, "In Search of the Winner's Curse," *Economic Inquiry*, October 1984, *22*, 579–92.

______, "In Search of the Winner's Curse: Reply," *Economic Inquiry*, July 1986, *24*, 517–20.

David, H. A., *Order Statistics*, 2nd ed., New York: Wiley & Sons, 1981.

Feinstein, Jonathan S., Block, Michael K. and Nold, Frederick C., "Asymmetric Information and Collusive Behavior in Auction Markets," *American Economic Review*, June 1985, *75*, 441–60.

Gilley, Otis W. and Karels, Gordon V., "The Competitive Effect in Bonus Bidding: New Evidence," *Bell Journal of Economics*, Autumn 1981, *12*, 637–48.

Gupta, A. K., "Estimation of the Mean and Standard Deviation of a Normal Population from a Censored Sample," *Biometrika*, 1952, *39*, 260–73.

Hausman, Jerry, "Simultaneous Equation Models," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Amsterdam: North-Holland, 1983, ch.7.

Hendricks, Kenneth, Porter, Robert H. and Boudreau, Bryan, "Information, Returns and Bidding Behavior in OCS Auctions: 1954–1969," *Journal of Industrial Economics*, June 1987, *35*, 517–42.

Johnson, Stanley R. and Kuhlman, John M., "The Number of Competitors and Bid Prices," *Southern Economic Journal*, July 1983, *50*, 213–20.

Kagel, John H. and Levin, Dan, "The Winner's Curse and Public Information in Common-Value Auctions," *American Economic Review*, December 1986, *76*, 894–920.

Kuhlman, John M. and Johnson, Stanley R., "Estimating Damages on Highway Construction Contracts," *Antitrust Bulletin*, Winter 1984, *29*, 719–38.

Lohrenz, John, "Relationships Between Bidding and Hydrocarbon Production of the Federal Outer Continental Shelf (through 1970)," U. S. Department of Commerce, National Technical Information Service No. PB-238-188, 1974.

McAfee, R. Preston and McMillan, John, "Auctions and Bidding," *Journal of Economic Literature*, June 1987, *25*, 699–754.

Milgrom, Paul R. and Weber, Robert J., "A Theory of Auctions and Competitive Bidding," *Econometrica*, September 1982, *50*, 1089–122.

Ramsey, James, "Empirical Analysis of Lease Bidding Using Historical Data," in R. Engelbrecht–Wiggans et al., eds., *Auctions, Bidding and Contracting: Uses and Theory*. New York: New York University Press, 1983, 285–306.

Stark, Robert M., "Unbalanced Highway Contract Tendering," *Operational Research Quarterly*, September 1974, *25*, 373–88.

______ **and Varley, Thomas C.,** "Bidding, Estimating, and Engineered Construction Contracting," in R. Engelbrecht–Wiggans et al., eds., *Auctions, Bidding and Contracting: Uses and Theory*. New York: New York University Press, 1983, 121–36.

Thaler, Richard, "The Winner's Curse," *Journal of Economic Perspectives*, Winter 1988, *2*, 191–202.

Wilson, Robert, "A Bidding Model of Perfect Competition," *Review of Economic Studies*, October 1977, *44*, 511–18.

______, "On Competitive Bidding Applied," in R. Engelbrecht-Wiggans et al., eds., *Auctions, Bidding and Contracting: Uses and Theory*. New York: New York University Press, 1983, 363–68.

Engineering News-Record, Weekly Issues, 1982–1985.

Learning in Equilibrium Models of Arbitration

By ROBERT GIBBONS*

This paper analyzes strategic communication in equilibrium models of conventional and final-offer interest arbitration. Both models emphasize the role of learning by the arbitrator from the parties' offers about the state of the employment relationship, which is known to the parties but not to the arbitrator. The analysis focuses on separating equilibria in which the arbitrator's behavior is identical to the reduced-form decision rule typically assumed in the empirical literature. The paper thereby provides a structural interpretation for the existing empirical work.

Unless an alternative is specified in advance, disputes in ongoing economic relationships must often be settled by drastic measures: breaking up the relationship or, perhaps, going to the courts. To avoid these substantial costs, an entire class of dispute-resolution techniques has been invented and is being promoted for use in many labor, commercial, and personal relationships.[1] One of the most interesting of these alternative dispute-resolution methods is binding arbitration, in which the parties agree that a third party will resolve their dispute.

Understanding arbitration is important for several reasons. First, arbitration plays an important and growing role in allocating resources, especially in wage disputes and in pre-trial arbitration. Second, arbitration offers a rare natural experiment involving strategic behavior and significant stakes: in some cases, the rules constitute a well-defined game and the players' choices are observed without error.

A third reason to understand arbitration is because it is an institution that allows parties to write an incomplete contract *ex ante* while limiting opportunistic behavior *ex post*. Such institutions are likely to be important in the emerging theories of internal organization and incomplete contracts. Oliver Williamson (1975), for instance, argues that the possibility of arbitration by informed insiders is one of the central virtues of internal over market organization. And Sanford Grossman and Oliver Hart (1986) analyze two other responses to incomplete contracts—integration (i.e., assigning residual rights of control) and *ex post* bargaining; arbitration is a third option.

The nature of the arbitrator's task depends on whether a contract is in force at the time of the dispute. In interest arbitration the arbitrator settles wage (or other) disputes that arise when a new contract is negotiated; in grievance arbitration, the arbitrator settles disputes that arise during the life of an existing contract. The two major forms of interest arbitration are conventional arbitration and final-offer arbitration. In final-offer arbitration, the parties simultaneously submit wage offers, and then the arbitrator imposes one of the two offers as the settlement; in conventional arbitration,

---

*Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139 and National Bureau of Economic Research. I am grateful for several forms of assistance and support: I have benefited greatly from extensive discussions about arbitration with David Bloom, Christopher Cavanagh, Henry Farber, and Stanley Jacks. Many others have made helpful comments on this paper: David Card, Vincent Crawford, Paul Klemperer, Gregory Leonard, Meg Meyer, Kevin M. Murphy, Joel Sobel, two anonymous referees, and seminar audiences at Chicago, Harvard, MIT, and Princeton. And the NSF provided financial support through grant IST 86-09691.

[1] Third-party intervention takes several forms and occurs in many contexts. On mediation, see Roger Fisher (1978) on international negotiations; Gerald Cormick and Leota Patton (1977) on environmental issues; and Thomas Kochan and Todd Jick (1978) on public-sector disputes. On arbitration, see Kochan (1980) on public- and private-sector labor disputes and Paul Joskow (1987) on commercial contracts in the coal industry.

the arbitrator's choice of a settlement is unconstrained.

This paper analyzes strategic communication in equilibrium models of both conventional and final-offer interest arbitration. Both models emphasize the role of learning by the arbitrator from the parties' offers about the state of the employment relationship, which is known to the parties but not to the arbitrator. In certain equilibria of both the conventional and the final-offer arbitration models, the arbitrator's preferred settlement, y_a, is a weighted average of the average of the parties' offers, $\bar{y} = (y_e + y_u)/2$, and a settlement motivated by the publicly observable facts of the case, $y^*(f)$:

$$(1) \qquad y_a = \alpha \bar{y} + (1 - \alpha) y^*(f).$$

A decision rule of this form also is common in empirical work on arbitration. It appears in studies of conventional arbitration as equation (12) in Orley Ashenfelter and David Bloom (1984), equation (3) in Henry Farber and Max Bazerman (1986), and equation (1b) in Bloom (1986), and it appears in studies of final-offer arbitration as equation (7) in Ashenfelter and Bloom and equation (6) in Farber and Bazerman. Thus, although this paper contains no data, it is closely related to the empirical literature on interest arbitration: the paper provides a structural interpretation for the estimated parameter α.

The paper also is an attempt to move toward a complete theory of interest arbitration. A first requirement for such a theory is that it both match the existing empirical evidence and provide new predictions for empirical testing. Perhaps the strongest piece of existing evidence is that the parties' offers matter in the arbitrator's decision making (i.e., α is positive; see Bloom and Farber-Bazerman). Indeed, Carl Stevens' (1966) proposal of final-offer arbitration was motivated by the observation that the offers seem to matter too much in conventional arbitration: the arbitrator sometimes seems simply to split the difference between the parties' offers, ignoring the publicly observable facts of the case. This paper uses the arbitrator's need to learn from the parties' offers to explain the observed importance of the offers in the arbitrator's decision making. The equilibrium behavior that emerges also is consistent with the Farber-Bazerman finding that the importance of the offers in the arbitrator's decision making increases when the offers are closer together, and with the Ashenfelter-Bloom and Farber-Bazerman finding that the arbitrator's notion of an ideal settlement is identical in final-offer and conventional arbitration.

A second requirement for a complete theory of arbitration is that it be an equilibrium analysis. The following discussion explains why an equilibrium analysis of arbitration is a crucial complement to empirical work on the arbitrator's decision rule. Consider Stevens' observation that in conventional arbitration the arbitrator sometimes seems simply to split the difference between the offers. If the arbitrator were committed to splitting the difference, then the parties would submit the most extreme offers possible. Suppose instead that the arbitrator is committed to splitting the difference between reasonable offers but to ignoring (or even punishing) unreasonable ones. Such a decision rule could cause the parties always to submit reasonable offers, and so lead to the appearance of mechanical compromising by the arbitrator, but the observed behavior would be only half the story: the data would be silent about how the arbitrator would have behaved had the parties not submitted reasonable offers. Out-of-equilibrium responses by the arbitrator are never observed, and so can never be estimated, but play a crucial role in determining what is observed; an equilibrium analysis makes these out-of-equilibrium responses explicit.

A third requirement for a complete theory of arbitration is that models of different forms of arbitration be built on a common set of assumptions about preferences, information, and commitment. This is a prerequisite for accurate comparisons of conventional and final-offer arbitration for two reasons, the first empirical and the second theoretical. First, if each form is adopted in environments in which it performs well (and if the data cannot reveal the differences in these environments), then comparisons of performance data from real arbitrations will suffer from selection bias. And second, because any form of arbitration could in principle be adopted in any given environment, a successful model of a given form of arbitration must be built on assumptions that permit other forms of arbitration to be analyzed within the same model. Consider Farber's (1980) model of final-offer arbitration, for example. Farber's model is the leading contribution to the existing theoretical literature on that form of arbitration, and is the basis for all of the empirical work on the subject. Its important liability, however, is that an analysis of conventional arbitration under Farber's assumptions about preferences, information, and commitment would imply that the parties' offers are irrelevant in the arbitrator's decision making, an implication that is strongly contradicted by the empirical evidence.

This paper satisfies these three requirements, but is far from a complete theory of arbitration. The paper's most important contributions are that it provides an equilibrium foundation for the intuition that the arbitrator might learn from the parties' offers, and that it uses the idea of learning to develop a unified analytical treatment of the two major forms of interest arbitration.

The remainder of the paper is organized as follows. Section I describes Farber's model of final-offer arbitration without learning. As will become clear, this model plays an important role in the development of the learning models of *both* final-offer and conventional arbitration analyzed here. Section II specifies the economic environment (i.e., makes assumptions about the players' preferences and information structures) that underlies the models of both final-offer and conventional arbitration. Sections III and IV analyze separating equilibria in the final-offer and conventional arbitration games, respectively. In both cases, the parties' offers perfectly reveal their private information, and the arbitrator uses this information to help determine an ideal settlement. In final-offer arbitration the arbitrator chooses the offer that is closer to this ideal as the settlement; in conventional arbitration the arbitrator simply imposes the ideal itself. The model of conventional arbitration in Section IV is of independent interest as a contribution to the theory of cheap-talk communication games. Section V summarizes the relationship between these learning models and the empirical literature. Finally, Section VI discusses a program of extensions of the models developed here.

I. Final-Offer Arbitration Without Learning

Farber characterizes an arbitrator by a number, z, that describes the arbitrator's most preferred settlement in the case in question. To facilitate comparison of Farber's model with those below, suppose that if the ideal settlement is z and the actual settlement is y then the arbitrator's utility is $v_a(y,z) = -(y-z)^2$. In final-offer arbitration, the arbitrator is constrained to choose one of the parties' offers as a settlement. Given this utility function, and assuming that the employer's offer, y_e, is below the union's offer, y_u, the arbitrator chooses the employer's offer if it is closer to the arbitrator's ideal settlement, that is, if and only if $z < \bar{y}$, where $\bar{y}$ is the average of the parties' offers.

The parties are assumed to be uncertain about the value of z. Let this uncertainty be described by the probability distribution $F(z)$, with density $f(z)$. Given the parties' offers, the probability that the employer's offer is accepted is then $F(\bar{y})$. Assume also that the parties have strictly opposed preferences and are risk-neutral: the employer simply seeks to minimize the arbitrator's expected settlement, while the union seeks to maximize it.

The timing of the final-offer arbitration game is as follows. First, the parties simultaneously submit their offers to the arbitrator. Then the arbitrator chooses the offer that maximizes the utility function $v_a(y,z)$. The parties' Nash equilibrium offers maximize their expected payoffs and so solve

$$(2) \quad \max_{y_u} y_e F(\bar{y}) + y_u[1 - F(\bar{y})], \quad \text{and}$$

$$(3) \quad \min_{y_e} y_e F(\bar{y}) + y_u[1 - F(\bar{y})].$$

The first-order conditions for these optimization problems are

$$(4)\qquad 1-F(\bar{y})=\tfrac{1}{2}(y_u-y_e)f(\bar{y}) \quad \text{and}$$

$$(5)\qquad F(\bar{y})=\tfrac{1}{2}(y_u-y_e)f(\bar{y}),$$

which imply that $F(\bar{y})=1/2$ and $y_u-y_e=f(\bar{y})^{-1}$. (Note that the assumption that $y_u>y_e$ clearly cannot be false in equilibrium.) These first-order conditions express a simple tradeoff: in equilibrium, each party considers making a more aggressive offer but must balance the gain from having such an offer accepted against the reduced probability that it will be accepted.

Much of what follows assumes that z is normally distributed with mean M and precision H (i.e., variance $1/H$). In this case, (4) and (5) imply that the equilibrium offers are

$$(6)\qquad y_u=M+\sqrt{\frac{\pi}{2H}} \quad \text{and}$$

$$(7)\qquad y_e=M-\sqrt{\frac{\pi}{2H}}\,.$$

(Checking for the global optimality of these and subsequent offers is tedious and so is relegated to Appendix I.) Note that the equilibrium offers are centered about the mean of the parties' belief about the arbitrator's ideal settlement, and that the distance between the equilibrium offers decreases as this belief becomes more precise.

II. An Economic Environment

This section describes the preferences and information structure that underlie both the final-offer and the conventional arbitration models analyzed in Sections III and IV. The novel feature introduced here is an unknown called the state of the employment relationship (or simply the state) and denoted by s. The most important distinction between the publicly observable facts of the case and the state of the employment relationship is simply that the arbitrator observes the former but not the latter. Examples of publicly observable facts are the inflation rate, recent wage settlements in related industries, a coarse indication of the firm's financial health, and so on. The state of the employment relationship, in contrast, represents the details of the production environment and the labor market that the employer and the union inhabit. Because of their long and close relationship, the employer and the union understand the state well; the arbitrator has less experience with these details, but conducts an independent investigation to gather information.

The theme of the models analyzed below is as follows. If the arbitrator knew the state precisely then the ideal settlement would be clear. Because the arbitrator is comparatively poorly informed, however, any information about the state that can be extracted from the parties' offers may be useful. Such inferences must of course account for the natural incentive of each party to mislead the arbitrator.

A. *Preferences*

Assume that the arbitrator's ideal settlement depends on the state of the employment relationship: $z=z(s)$. This dependence could arise because the arbitrator wishes to be fair, or to be rehired. For simplicity, let $z(s)=s$ and let the arbitrator's (von Neumann-Morgenstern) utility function be $v_a(y,s)=-(y-s)^2$.

Unlike the arbitrator, the parties are assumed to be risk-neutral and to have preferences that are independent of the state: $v_e(y,s)=-y$ and $v_u(y,s)=y$. As in the previous section, the employer simply seeks to minimize the arbitrator's expected settlement, while the union seeks to maximize it. Thus, the parties are well informed about what a fair settlement would be, but this information does not affect their preferences.

These assumptions about preferences keep the analysis simple. It may seem odd, however, that the state appears in the arbitrator's utility function but not in the parties' utility functions. Two justifications for this can be given. First, although new issues will arise, the basic incentives analyzed here seem unlikely to disappear when the parties' preferences do depend on the state. And second, the assumptions could be correct as they stand: Let the value of the enterprise be $V(s)$ when the state is s, let the settlement y be the wage bill, and let employment be fixed. Then the union's utility function is $v_u(y,s)=y$ and the employer's is $v_e(y,s)=V(s)-y$. The presence of $V(s)$ in the employer's utility function changes nothing in the analysis, and the arbitrator's ideal settlement $z(s)=s$ can be interpreted as a judgment about the fair division of $V(s)$.

B. *Information Structure*

Let s be normally distributed with mean m and precision h. Also, let the parties' knowledge of the state be summarized by the noisy signal

$$s_p=s+\varepsilon_p, \tag{8}$$

where ε_p is normally distributed with zero mean and precision h_p. Notice that the parties are assumed to share a common perception of the state of the employment relationship. While it seems realistic that the parties should both be considerably better informed about the state than the arbitrator is, it also seems likely that the parties themselves will have different beliefs about the state. This assumption is important in what follows, and is discussed in Section VI.

Let the results of the arbitrator's private investigation be represented by the noisy signal

$$s_a=s+\varepsilon_a, \tag{9}$$

where ε_a is normally distributed with zero mean and precision h_a. The three random variables s, ε_p, and ε_a are assumed to be independent of each other. In keeping with the idea that the parties are better informed about the state than is the arbitrator, one could assume that $h_p>h_a$, but this is not necessary for what follows.

To summarize, the information structure is as follows: the parties both observe s_p, the arbitrator observes s_a, no one observes s, and m, h, h_p, and h_a are common knowledge. The parameters m and h that specify the prior distribution of s could reasonably be interpreted as the publicly observable facts of the case.

Because of the independence of s, ε_p, and ε_a, the simple formulas of the normal learning model (Morris DeGroot, 1970, ch. 9) characterize the Bayesian updating that occurs in the model. Some of the results of this updating are recorded here for future use. The conditional distribution of s given s_i (where $i\in\{p,a\}$) is normal with mean $M_i(s_i)$ and precision H_i, where

$$M_i(s_i)=\frac{hm+h_is_i}{h+h_i} \quad \text{and} \tag{10}$$

$$H_i=h+h_i. \tag{11}$$

Similarly, the conditional distribution of s given s_p and s_a is normal with mean $M_{pa}(s_p,s_a)$ and precision H_{pa}, where

$$M_{pa}(s_p,s_a)=\frac{hm+h_ps_p+h_as_a}{h+h_p+h_a} \tag{12}$$

and

$$H_{pa}=h+h_p+h_a. \tag{13}$$

These definitions provide a convenient shorthand for describing the beliefs and decision making in the models below.

III. Learning in Final-Offer Arbitration

This section analyzes a model of final-offer arbitration in the economic environment specified in Section II. The analysis focuses on a separating equilibrium. The arbitrator perfectly infers the parties' private information (s_p) from their offers and then uses this information and the signal (s_a) to compute a posterior belief about the state (s) and to choose the expected utility-maximizing offer as the settlement. The parties understand that the arbitrator will draw an inference from the offers and so consider the payoff from misleading the arbitrator when choosing their offers, but find it optimal not to submit misleading offers.

The timing of the game is as follows. First, the parties observe s_p and the arbitrator observes s_a. Second, the parties simultaneously submit offers, y_e and y_u. Third, the arbitrator chooses the offer that maximizes $E_s[v_a(y,s)|s_a, y_e, y_u]$, where the conditioning on y_e and y_u denotes the information about s_p that the arbitrator extracts from the parties' offers. Note that the arbitrator is assumed to be unable to commit to a decision rule before the parties submit their offers; rather, the arbitrator's choice must satisfy sequential rationality at the third stage of the game. Taken literally, this assumption seems to represent observed practice accurately, but the complementary case also deserves attention; see the discussion in Section VI.

Before describing the separating equilibrium in detail, it helps set the tone for the subsequent analysis to show that there does *not* exist an equilibrium in which the arbitrator simply ignores the parties' offers. More precisely, there is no perfect Bayesian equilibrium in which the arbitrator's belief is the conditional distribution of s given s_a both on and off the equilibrium path.[2] To see that this is so, suppose the contrary. Then the arbitrator's ideal settlement, $y_a(s_a)$, would be $M_a(s_a)$ from (10), or

$$(14)\qquad y_a(s_a)=\frac{hm+h_a s_a}{h+h_a}.$$

Given s_p, the parties' belief about $y_a(s_a)$ is normally distributed with mean m' and precision h', where

$$(15)\qquad m'=\frac{hm+h_a M_p(s_p)}{h+h_a}\quad\text{and}$$

$$(16)\qquad h'=\frac{(h+h_p)(h+h_a)^2}{h_a(h+h_p+h_a)}.$$

Note that the dependence of m' on s_p has been suppressed in the notation.

The logic behind Farber's equilibrium now applies directly: if the arbitrator's belief is independent of the parties' offers, then s_p affects the parties' offers only because it changes their belief about the arbitrator's ideal settlement. The equilibrium offers are therefore analogous to (6) and (7):

$$(17)\qquad y_u(s_p)=m'+\sqrt{\frac{\pi}{2h'}}\quad\text{and}$$

$$(18)\qquad y_e(s_p)=m'-\sqrt{\frac{\pi}{2h'}}.$$

The key point is that the arbitrator *can* learn something from these offers. The average of the offers is m', and (10) and (15) determine s_p from m'. Let $s_p(m')$ denote the (point) estimate of s_p computed in this way. Sequential rationality then requires that the arbitrator's ideal settlement be $M_{pa}(s_p(m'), s_a)$ from (12), or

$$(19)\qquad y_a(y_e, y_u, s_a)=\frac{hm+h_p s_p(m')+h_a s_a}{h+h_p+h_a}.$$

The crucial difference between (14) and (19), of course, is that in the latter the parties' offers help the arbitrator determine the ideal settlement. In an equilibrium featuring this kind of learning, each party's offer must balance not only the probability of and the gain from submitting the offer chosen as the settlement, as in Farber's model, but also the effect the offer has on the arbitrator's inference about the ideal settlement. The rest of this section establishes that there exists a separating perfect Bayesian equilibrium in this final-offer arbitration game.

---

[2] There is, however, a continuum of pooling equilibria in which the arbitrator learns nothing from the parties' offers in equilibrium, but does learn something off the equilibrium path, in the sense that the arbitrator's out-of-equilibrium belief is not the conditional distribution of s given s_a. Suppose, for instance, that for every realization of s_p the employer's offer is $\hat{y}$ and the union's offer is $\tilde{y}$. The values of $\hat{y}$ and $\tilde{y}$ are arbitrary, except that $\hat{y}$ must not exceed $\tilde{y}$. Note that the expected settlement if the parties follow these pooling offer strategies is a weighted average of $\hat{y}$ and $\tilde{y}$, with probability weights that depend on the realization of s_p. Nonetheless, it is simple to construct an out-of-equilibrium belief for the arbitrator that rationalizes out-of-equilibrium settlements that deter each party from deviating, *independent* of the realization of s_p. If the employer deviates to y', for instance, then the arbitrator's belief can rationalize choosing the union's offer ($\tilde{y}$) when $y' < \tilde{y}$ and can rationalize choosing the employer's offer (y') when $y' \geq \tilde{y}$.

Suppose that the arbitrator believes that $\bar{y}$, the average of the parties' offers, perfectly reveals s_p, both on and off the equilibrium path. That is, for *any* pair of offers, y_e and y_u, the arbitrator computes the point estimate $s_p = s_p(\bar{y})$. (Other separating equilibria may exist in which a different function of y_e and y_u reveals s_p to the arbitrator.) The arbitrator's ideal settlement is then

$$(20)\quad y_a(y_e, y_u, s_a) = \frac{hm + h_p s_p(\bar{y}) + h_a s_a}{h + h_p + h_a},$$

and (assuming $y_e < y_u$) the arbitrator chooses y_e as the settlement if and only if the ideal settlement, $y_a(y_e, y_u, s_a)$, is less than the average of the offers, $\bar{y}$.

If the equilibrium value of $\bar{y}$ is to reveal s_p in this way then the arbitrator's inference rule must be

$$(21)\quad s_p(\bar{y}) = \frac{(h + h_p)\bar{y} - hm}{h_p}.$$

To see this, use (20) to express the event that the arbitrator chooses y_e as $s_a < S(\bar{y})$, where

$$(22)\quad S(\bar{y}) = \frac{h_a\bar{y} + h(\bar{y} - m) + h_p(\bar{y} - s_p(\bar{y}))}{h_a}.$$

A derivation analogous to that leading to (4) and (5) then yields the first-order conditions

$$(23)\quad 1 - F[S(\bar{y})|s_p] = (y_u - y_e) f[S(\bar{y})|s_p] S'(\bar{y})$$

and

$$(24)\quad F[S(\bar{y})|s_p] = (y_u - y_e) f[S(\bar{y})|s_p] S'(\bar{y}),$$

where F is now the distribution of s_a conditional on s_p, which is normal with mean $M_p(s_p)$ given by (10) and precision

$$(25)\quad H' = \frac{h_a(h + h_p)}{h + h_p + h_a}.$$

As in (6) and (7), these first-order conditions imply that $S(\bar{y}) = M_p(s_p)$. Substituting $s_p(\bar{y})$ for s_p in $M_p(s_p)$ and solving for $s_p(\bar{y})$ yields (21), the desired result.

The inference rule $s_p(\bar{y})$ in (21) implies that (22) simplifies to $S(\bar{y}) = \bar{y}$: the arbitrator's decision rule is to choose y_e if and only if $s_a < \bar{y}$. Note that this is directly analogous to the decision rule used in Farber's model (where there is no learning). Not surprisingly, therefore, the equilibrium strategies for the parties that solve the first-order conditions (23) and (24) are directly analogous to (6) and (7):

$$(26)\quad y_u(s_p) = M_p(s_p) + \sqrt{\frac{\pi}{2H'}} \quad \text{and}$$

$$(27)\quad y_e(s_p) = M_p(s_p) - \sqrt{\frac{\pi}{2H'}}.$$

The average of the offers is $M_p(s_p)$, which yields a more convenient expression of the inference rule (21): in equilibrium, the arbitrator infers s_p from the average of the parties' offers, uses it in (20) to compute the ideal settlement, and then chooses the offer closer to this ideal.

The summary of this argument is the main result of this section.

PROPOSITION 1: *The parties' offer strategies given in* (26) *and* (27) *and the arbitrator's decision strategy based on the ideal settlement* (20) *and the inference rule* (21) *constitute a separating perfect Bayesian equilibrium of the final-offer arbitration game. In this equilibrium, the arbitrator's ideal settlement can be written*

$$y_a = \alpha\bar{y} + (1 - \alpha)s_a, \quad \text{where}$$

$$\bar{y} = \frac{hm + h_p s_p}{h + h_p} \quad \text{and}$$

$$\alpha = \frac{h + h_p}{h + h_p + h_a}.$$

The key feature of the arbitrator's equilibrium strategy is its treatment of unexpected offers from the parties. In equilibrium, the parties' offers differ by $\sqrt{2\pi/H'}$, and the arbitrator perfectly infers the parties' private information from the average of the offers. Any pair of offers that differs by this prescribed amount is on the equilibrium path; any pair that differs by any other amount is off the equilibrium path. The arbitrator's inference rule ignores this distinction, however: (21) produces the point estimate of s_p both on and off the equilibrium path. Thus, any change in either party's offer changes the arbitrator's estimate of s_p at the rate $(h+h_p)/2h_p$. The reason the arbitrator is not misled in equilibrium is that the parties must balance this opportunity to influence the arbitrator's belief against the two considerations that determined the equilibrium offers in Farber's model, the gain from having a more aggressive offer accepted and the reduced probability that it will be accepted.

There are two cases in which learning plays no role in the private-information game analyzed here: (1) the arbitrator has nothing to learn (i.e., h_a approaches infinity); and (2) the parties have nothing to communicate (i.e., h_p approaches zero). One appealing feature of the separating equilibrium described in Proposition 1 is that it approaches the equilibrium in Farber's complete-information game as the learning becomes unimportant in the private-information game. As h_a approaches infinity, for instance, the arbitrator's uncertainty about s disappears, and s_p becomes relevant only as a signal to the parties about the arbitrator's ideal settlement. In this case, the precision of the parties' belief about the arbitrator's ideal settlement (i.e., H in (6) and (7) in Farber's model) is $h+h_p$, which is indeed the limit of H' in (26) and (27) as h_a approaches infinity. Similarly, as h_p approaches zero, the parties' private information does not improve their prior information about the arbitrator's ideal settlement, so H in (6) and (7) is $hh_a/(h+h_a)$, which is indeed the limit of H' as h_p approaches zero.[3]

IV. Learning in Conventional Arbitration

This section analyzes a model of conventional arbitration in the economic environment specified in Section II. The analysis focuses on a continuum of payoff-equivalent separating equilibria. In each of these equilibria, the arbitrator perfectly infers the parties' private information (s_p) from their offers and then uses this information and the signal (s_a) to compute a posterior belief about the state (s) and to impose whatever settlement maximizes expected utility given this belief. The parties understand that the arbitrator will draw an inference from the offers and so consider the payoff from misleading the arbitrator when choosing their offers, but find it optimal not to submit misleading offers.

The timing of the game is as follows. First, the parties observe s_p and the arbitrator observes s_a. Second, the parties simultaneously submit offers, y_e and y_u. Third, the arbitrator imposes whatever settlement maximizes $E_s[v_a(y,s)|s_a, y_e, y_u]$. As in the previous section, the arbitrator is assumed to be unable to commit to a decision rule in advance. It will become clear below, however, that in this model of conventional arbitration the arbitrator would not value the opportunity to commit to a decision rule: each of the continuum of sequentially rational decision rules associated with the continuum of separating equilibria described below would be an optimal rule if the arbitrator could commit.

---

[3] One other limit also bears consideration. As h_p approaches infinity, the arbitrator perfectly infers not only s_p but also s from the average of the parties' offers. Given perfect information about s, the arbitrator's private signal s_a becomes irrelevant in determining the arbitrator's ideal settlement, and this eliminates the parties' uncertainty about the arbitrator's ideal settlement. Recall that as the analogous uncertainty disappears in Farber's complete-information model, the parties' equilibrium offers converge to the mean of the distribution of the arbitrator's ideal settlement. (Formally, as H approaches infinity, the equilibrium offers given in (6) and (7) converge to M.) Interestingly, as h_p approaches infinity, H' does not approach infinity (rather, it approaches h_a), so the equilibrium offers given by (26) and (27) in the private-information model do not converge. This happens because (as noted above) (21) and (22) imply that the arbitrator chooses y_e if and only if $s_a < \bar{y}$, independent of the value of h_p. Thus, as h_p approaches infinity, s_a serves as a tie breaker in the arbitrator's decision rule even though it is irrelevant to the arbitrator's inference process.

The crucial difference between this model of conventional arbitration and the previous section's model of final-offer arbitration is that here arbitration is modeled as a *cheap-talk* game: the parties' offers are not directly relevant to any player's payoff; they matter only indirectly through the information they convey.[4] In this case, it is natural in the model (as well as faithful to the institutional details of arbitration) to envision party i's offer as a long speech that ends "and that is why an appropriate settlement is y_i." Final-offer arbitration is not a cheap-talk game because one of the parties' offers must be chosen as the settlement, and so directly affects payoffs. Similarly, conventional arbitration would not be a cheap-talk game if the offers in the arbitration game were constrained to be the parties' last offers in a prior bargaining game, because such offers could have been accepted and would then have directly affected payoffs.[5] An important problem for future research will be to embed the arbitrator's learning problem studied here in a two-stage model that begins with a bargaining stage and then proceeds to an arbitration stage if the parties do not negotiate a settlement in the first stage.

All cheap-talk games share two distinctive features. The first is that in any cheap-talk game there exists a trivial equilibrium in which no communication occurs. In such an equilibrium in the conventional arbitration model, the arbitrator expects there to be no information in either party's offer and so pays no attention to either offer when choosing a settlement, and each party is content to choose an offer at random because no offer has any effect on the arbitrator's settlement. Note well, however, that in this equilibrium the arbitrator's ideal (and imposed) settlement is independent of the parties' offers. Since such independence is not observed in the data, this no-communication equilibrium is suspect.

The second distinctive feature of all cheap-talk games is that, given a fixed cheap-talk game, any perfect Bayesian equilibrium (separating or otherwise) is associated with a class of payoff-equivalent perfect Bayesian equilibria that differ only by a translation: since messages are costless, the information communicated in any given equilibrium also can be communicated in other payoff-equivalent equilibria that differ only in the language used to communicate. This accounts for the great multiplicity of payoff-equivalent separating equilibria mentioned above.

This section's model of conventional arbitration is a cheap-talk game with a third distinctive feature: neither party alone can achieve any communication with the arbitrator, because each party's preferences over settlements are independent of the state of the employment relationship. (The employer simply seeks to minimize the arbitrator's expected settlement, while the union seeks to maximize it.) This section demonstrates, however, that together the parties can achieve perfect communication: There exist separating perfect Bayesian equilibria in which the parties' offers perfectly reveal s_p to the arbitrator.

One method for extracting a single piece of information from two informed agents is to ask each to make a claim about the information and to shoot them both if their claims disagree.[6] This method is not feasible in the arbitration model analyzed here, however, for two reasons. First, the arbitrator's power is limited to imposing a transfer between two risk-neutral parties, so outcomes that are Pareto-inefficient for the two parties are impossible. And second, the arbitrator's behavior must be sequentially rational: given any pair of offers (y_e, y_u) and any signal (s_a), the arbitrator must formulate a belief about the parties' private information, $\mu(s_p|s_a, y_e, y_u)$, and act optimally given this belief.

---

[4] See Vincent Crawford and Joel Sobel (1982) for the seminal work on cheap-talk games.

[5] Thus, in this paper, the parties' offers in arbitration need not be identical to their last offers in a prior bargaining game. This flexibility arises if (as is required in some contracts) the bargaining record has been sealed and is unavailable to the arbitrator. Also, it will become clear below that the incentive for the parties to make concessions in arbitration can be quite strong, perhaps stronger than the analogous incentive in bargaining. Therefore, even if the bargaining record has not been sealed, the last offers from the bargaining game may be sufficiently far apart that they have no bearing on the parties' choices of offers in arbitration.

[6] See Ehud Kalai and Robert Rosenthal (1979) for a more careful (if less dramatic) statement of this result.

In spite of the facts that Pareto-inefficient outcomes are impossible and that the arbitrator must be sequentially rational, the arbitrator can extract s_p from the parties. The crucial feature of the model that drives this result is that the parties' information (s_p) is correlated with the signal from the arbitrator's investigation (s_a), because both are noisy estimates of the underlying state of the employment relationship (s).

To understand this result, it is helpful to consider the intermediate case in which Pareto-inefficient outcomes are impossible but sequential rationality is not required. In this case, the following scheme induces truth-telling: ask each party for a claim about s_p; if they disagree, then compare each party's claim about s_p to the signal s_a and enforce a large transfer to the party who seems to have made the more reasonable claim from the party who seems to have made the less reasonable claim. Given any correlation between s_p and s_a, there exists a sufficiently large transfer such that truth-telling is the unique equilibrium of this scheme. Furthermore, this scheme may not be unreasonable in practice: The arbitrator gives both parties the opportunity to present their cases fully and to challenge assertions in the other's case. Afterward, when all the cards are supposed to be on the table, the arbitrator asks the parties whether they can now resolve their dispute. If they cannot, then the arbitrator (thinking that reasonable people should by this point agree) conducts an investigation, compares the result to the parties' claims, and throws the book at the party who seems to have made the less reasonable claim.

It is simple to extend the spirit of this scheme to the model of conventional arbitration studied here, in which the arbitrator's behavior must be sequentially rational. The argument relies on the fact that in a perfect Bayesian equilibrium of the conventional arbitration game, the arbitrator's belief off the equilibrium path is unrestricted. This freedom ensures that there exists a belief off the equilibrium path that makes it sequentially rational for the arbitrator to enforce the required transfer if the parties disagree.

There is a close relationship between the incentives created in a conventional arbitration game in which the arbitrator uses the scheme described above and the incentives created in a final-offer arbitration game: the size of the transfer and the conditions under which it is awarded to one party or the other in the conventional arbitration game are analogous to the gap between the offers and the conditions under which the arbitrator chooses one party's offer or the other's in the final-offer arbitration game. In fact, the proof that there exists a continuum of separating perfect Bayesian equilibria in the conventional arbitration game with learning relies on the construction of the following continuum of final-offer games *without* learning.

Recall that in Section I the arbitrator's ideal settlement was denoted by z, and the parties' shared belief about z was denoted by $F(z)$. Suppose instead that for some fixed $k \in [0,1]$ the arbitrator's ideal settlement is

$$(28) \quad z(s_a|k) = \frac{(h+h_p)s_a - khm}{(1-k)h+h_p}.$$

Note that if the ideal settlement is given by (28) then there is nothing the arbitrator needs to learn from the parties: s_a completely determines the arbitrator's ideal settlement, independent of s_p. In this case, the only role for s_p is to help the parties forecast $z(s_a|k)$, just as in the derivation of (17) and (18) above. Given s_p, the parties' shared belief about $z(s_a|k)$ is normal with mean $M(s_p|k)$ and precision H_k, where

$$(29) \quad M(s_p|k) = \frac{(1-k)hm + h_p s_p}{(1-k)h+h_p} \quad \text{and}$$

$$(30) \quad H_k = \frac{h_a[(1-k)h+h_p]^2}{(h+h_p)(h+h_p+h_a)}.$$

From (6) and (7), the parties' equilibrium offers in this model of final-offer arbitration without learning are

$$(31) \quad y_u(s_p|k) = M(s_p|k) + \sqrt{\frac{\pi}{2H_k}} \quad \text{and}$$

$$(32) \quad y_e(s_p|k) = M(s_p|k) - \sqrt{\frac{\pi}{2H_k}}.$$

These equilibrium offers from the final-offer game without learning will turn out to be identical to the equilibrium offers in this section's model of conventional arbitration with learning!

Imagine that in the model of conventional arbitration with learning, the arbitrator believed that the parties' offer strategies were (31) and (32). The arbitrator could then invert the offer strategies so as to interpret the offers as implicit claims about the parties' information. Given the offer y_u from the union, for instance, the arbitrator could solve (31) for s_p to arrive at the implicit claim that the parties' private information is

$$(33) \quad s_{pu}(y_u|k) = \sigma\left(y_u - \sqrt{\pi/2H_k}|k\right),$$

where

$$(34) \quad \sigma(x|k) = x + \frac{(1-k)h}{h_p}(x - m)$$

solves $M(s_p|k) = x$ for s_p. Similarly, given the offer y_e from the employer, the arbitrator could solve (32) for s_p to arrive at the implicit claim that the parties' private information is

$$(35) \quad s_{pe}(y_e|k) = \sigma\left(y_u + \sqrt{\pi/2H_k}|k\right).$$

Note that if the parties' offers are in fact determined by (31) and (32) then these implicit claims will be identical even though the offers are not. If the implicit claims agree, as they will in equilibrium, then it is natural for the arbitrator to believe the common claim. Everything hinges, however, on what the arbitrator believes when the implicit claims disagree.

Suppose that after observing the signal s_a and the offers y_e and y_u the arbitrator arrives at the point estimate $s_p = s_p(y_e, y_u, s_a)$. Then sequential rationality dictates that the arbitrator impose the ideal settlement $M_{pa}[s_p(y_e, y_u, s_a), s_a]$ from (12), or

$$(36) \quad y_a(y_e, y_u, s_a) = \frac{hm + h_p s_p(y_e, y_u, s_a) + h_a s_a}{h + h_p + h_a}.$$

It turns out to be more convenient to express the arbitrator's inference rule not as $s_p(y_e, y_u, s_a)$, but as $M(y_e, y_u, s_a|k)$, where (29) yields

$$(37) \quad M(y_e, y_u, s_a|k) = \frac{(1-k)hm + h_p s_p(y_e, y_u, s_a)}{(1-k)h + h_p}.$$

This last piece of notation makes it possible to state the section's main result in a compact form. (The proof of the Proposition is given in Appendix II.)

PROPOSITION 2: *For each* $k \in [0,1]$ *there exists a separating perfect Bayesian equilibrium of the conventional arbitration game. In each of these equilibria, the arbitrator's ideal* (*and imposed*) *settlement is* (36). *For each fixed k, the parties' offer strategies are* (31) *and* (32), *and the arbitrator's inference rule is*

$$M(y_e, y_u, s_a|k) = \begin{cases} y_u & \text{if } s_{pe}(y_e|k) \neq s_{pu}(y_u|k) \\ & \quad \text{and } z(s_a|k) > \bar{y}, \\ \bar{y} & \text{if } s_{pe}(y_e|k) = s_{pu}(y_u|k), \\ y_e & \text{if } s_{pe}(y_e|k) \neq s_{pu}(y_u|k) \\ & \quad \text{and } z(s_a|k) \leq \bar{y}. \end{cases}$$

Given k, the arbitrator's ideal settlement can be written

$$y_a = \alpha_k \bar{y}_k + (1-\alpha_k)\frac{khm + h_a s_a}{kh + h_a}, \quad \text{where}$$

$$\bar{y}_k = \frac{(1-k)hm + h_p s_p}{(1-k)h + h_p} \quad \text{and}$$

$$\alpha_k = \frac{(1-k)h + h_p}{h + h_p + h_a}.$$

Several observations about this plethora of equilibria are in order. First, as noted earlier, all of the equilibria described in the Proposition are payoff-equivalent: even though changes in k affect both the mean of and the distance between the parties' offers, differences in k are simply differences in the languages used to communicate.[7] Second, two values of k are simple to interpret: $k = 0$ and $k = 1$. When $k = 0$, (29) becomes $M(s_p|k) = M_p(s_p)$, so in equilibrium the arbitrator interprets $\bar{y}$ as the parties' expectation of s given s_p. When $k = 1$, (29) becomes $M(s_p|k) = s_p$, so in equilibrium the arbitrator interprets $\bar{y}$ as s_p itself. Third, in the separating equilibrium with $k = 0$ the parties' offers in conventional arbitration are identical to their offers in the separating equilibrium in final-offer arbitration described in Section III. And fourth, in each of these separating equilibria the arbitrator has themaximum amount of information in imposing a settlement. Thus, for any k, the decision rule described in Proposition 2 would be an optimal rule if the arbitrator could commit.

V. Relationship of the Learning Models to the Empirical Literature

This section describes the connections between the empirical literature on arbitration and the separating equilibria of the learning models developed in Sections III and IV. As described above, much of the empirical work on arbitrator behavior assumes that the arbitrator's ideal settlement is given by the weighted average in (1). Propositions 1 and 2 describe equilibria in which the arbitrator's behavior takes exactly this form. This paper thus provides a structural interpretation for the empirical literature's reduced-form model: in both conventional and final-offer arbitration, the arbitrator's ideal settlement depends on the parties' offers because the offers convey information, and the importance of the offers (as measured by α) increases as the importance of this information increases relative to the publicly observable information and the arbitrator's independent investigation (i.e., as h_p increases relative to h and h_a).

The existing empirical literature also has generated several other stylized facts, beyond the significance of α in (1). One such stylized fact is that the arbitrator's ideal settlement is identical in the two forms of arbitration; see Ashenfelter-Bloom and Farber-Bazerman. The separating equilibria in the learning models analyzed here are founded on just such an invariant notion of an ideal settlement: under *any* form of arbitration, if the arbitrator perfectly infers s_p from the parties' offers then the ideal settlement is $M_{pa}(s_p, s_a)$ given in (12).

A second stylized fact is that the parties' offers are more important in the arbitrator's decision making (i.e., α is higher) when they are closer together; see Farber-Bazerman. The separating equilibria in both Sections III and IV predict that α should be higher when the offers are closer together, but suggest that this need not be a causal relationship. Consider Proposition 1, for instance. Note that α increases with h_p. In equilibrium, the difference between the offers in (26) and (27) is $\sqrt{2\pi/H'}$, where H' is given by (25) and also increases with h_p. Thus, an increase in h_p results in both an increase in α and a decrease in the difference between the offers, thereby producing the observed correlation.[8] The causal story behind the correlation is as follows. The arbitrator puts more weight on the parties' offers when the parties' information is known to be more precise, and hence more valuable in the arbitrator's inference process. At the same time, when the parties' information is more precise, the parties' belief about the arbitrator's private signal, s_a, is more precise, and this causes the parties to submit offers that are closer together.

---

[7] In fact, the argument above holds for any $k \in (-\infty,\infty)$ except for $k = h + h_p/h$, although extreme values of k imply that α_k is outside $[0,1]$.

[8] The precise relationship between α and $(y_u - y_e)$ is $\alpha = 2\pi/h_a(y_u - y_e)^2$. Apparently the parameter h_a can be identified econometrically from this relationship.

The model of final-offer arbitration developed in Section III does *not* correspond well to one other stylized fact: Ashenfelter and Bloom find that the union's offer is selected as the settlement more than half the time in New Jersey police-officer disputes, while Ashenfelter, James Dow, and Daniel Gallagher (1986) find that the union's offer is selected less than half the time in Iowa public-sector disputes; in Section III, the parties' offers are equally likely to be selected. Farber's model of final-offer arbitration without learning allows the parties to differ in their risk aversion. The more risk-averse party prefers to submit a conservative offer with a large probability of being chosen rather than an aggressive offer with a small probability. It would be interesting to know whether a learning model with parties of differing risk aversion could integrate these empirical findings on offer-selection frequencies with the role of the parties' offers in the arbitrator's inference process emphasized here.

VI. Extensions

Although this paper satisfies three of the conditions that will be required of a complete theory of arbitration, much work remains to be done. This section is limited to describing two extensions of the models analyzed above and to providing further discussion of the paper's two most important assumptions.

A. *Comparing Conventional and Final-Offer Arbitration*

This paper analyzes conventional and final-offer arbitration within a single economic environment. The obvious next step is a welfare comparison of these and other forms of arbitration. Unfortunately, the model analyzed here is not rich enough to make the comparison interesting. First, because the parties are risk-neutral and have strictly opposed preferences, any preference by one party for one form of arbitration would be exactly reversed for the other party. And second, the arbitrator clearly cannot improve upon the separating equilibrium in conventional arbitration, where the ideal settlement is based on full information and the actual settlement is unconstrained.

A more interesting comparison might arise if the parties were risk-averse. If a separating equilibrium exists in conventional arbitration when the parties are risk-averse then this will again be the arbitrator's preferred form of arbitration. The parties, however, might find that final-offer arbitration offers better insurance than conventional arbitration does. If the parties are allowed to choose the form of arbitration then the arbitrator's preference for conventional arbitration may be overridden.

B. *Modeling Other Forms of Arbitration*

Ashenfelter, Dow, and Gallagher study an interesting variant of final-offer arbitration that is practiced in Iowa. The rules are as follows: first, an arbitrator, acting as a fact finder, suggests a settlement; then the parties either accept the settlement or continue their dispute and submit offers; finally, if the dispute continues then a second arbitrator (drawn from the same pool as the first) chooses one of the parties' offers *or* the fact finder's proposed settlement.

It is straightforward to extend the model described in Section III in order to analyze this version of final-offer arbitration. Suppose the fact finder's investigation yields the signal $s_f = s + \varepsilon_f$, where ε_f is normal with zero mean and precision h_f and is independent of the other random variables in the model. Then (as in the conventional-arbitration model of Section IV) the fact finder's proposed settlement should be $y_f(s_f) = M_f(s_f)$, given by (10). The rest of the model is as before, except for two changes. First, both parties and the second arbitrator update their beliefs after observing the fact finder's proposal. And second, the second arbitrator now has a third choice in the feasible set of imposed settlements.

Ashenfelter, Dow, and Gallagher (1986) argue that the "extent to which the second arbitrator does not concur in the fact finder's proposal is an obvious measure of the degree of arbitral uncertainty that exists in the system" (p. 2). While this interpretation is certainly plausible, the disagreement between arbitrators also may reflect the second arbitrator's learning from the parties' offers.

C. *Commitment by the Arbitrator*

One of the important assumptions in this paper is that the arbitrator is unable to commit *ex ante* to a decision rule, even though this might induce a socially desirable outcome of the negotiation and arbitration process. This assumption seems to represent observed practice accurately: arbitrators do not commit to decision rules before the parties make their offers. On the other hand, the assumption ignores the important possibility that an arbitrator might develop a reputation in the course of repeated play. The arbitrator-selection mechanisms described and analyzed by Bloom and Christopher Cavanagh (1986a, b) will play an important role in the eventual analysis of such dynamic games. Because the complete dynamic analysis is likely to be difficult, however, it would be worthwhile to analyze the models that complement those in this paper—static models in which the arbitrator is allowed to commit to a decision rule *ex ante*. As noted in Section IV, each of the continuum of decision rules described there is an optimal rule if the arbitrator can commit in conventional arbitration. The nature of the optimal rule under commitment in final-offer arbitration is an open question.

D. *Arbitration Between Parties with Symmetric Information*

The second important assumption in this paper is that the parties have symmetric information about the state of the employment relationship. It seems likely, instead, that each of the parties will have private information about some aspects of the employment relationship. Such an information asymmetry between the parties may influence both the means and the substance of the parties' equilibrium communication with the arbitrator; this is another direction for further work. Perhaps more importantly, it prompts the question of how the parties came to be in arbitration.

The theory of bargaining under symmetric information, due to John Nash (1950) and Ariel Rubinstein (1982), suggests that negotiations should not fail when information is symmetric, especially when the arbitration game that follows failed negotiations imposes costs (such as the arbitrator's fee) on the parties. This does not imply, however, that the prospect of arbitration has no effect. Crawford (1982a), for instance, analyzes the way arbitration changes the threat point in a Nash-bargaining game: arbitration has an effect even though the bargainers never fail to settle. Crawford's analysis involves models of conventional and final-offer arbitration without learning; it would be interesting to know what new effects arise when learning is introduced. For example, when the realization of the state is extreme, and so an appropriate settlement is unattractive to one party, the fact that learning links the arbitrated settlement to the state implies that the disadvantaged party cannot escape from an unattractive negotiated settlement through arbitration by an uninformed arbitrator.

Thus, the prospect of arbitration between parties with symmetric information can have an effect even if the parties never enter arbitration. Furthermore, in richer models than those developed by Nash and Rubinstein, disputes *can* occur between parties with symmetric information. Crawford (1982b), for instance, formalizes Thomas Schelling's (1956) view that bargaining is a struggle between the bargainers to commit themselves to favorable positions. Attempts at commitment are uncertain to succeed, but are irreversible if they do. If only one party's attempt succeeds then that party wins a favorable negotiated settlement, but if the parties are both committed to incompatible positions then negotiations fail and arbitration may occur. A model based on these ideas can produce disputes between parties with symmetric information, but more work needs to be done in modeling the process of commitment.

One intriguing approach to modeling the process of commitment emphasizes the idea that the parties are groups rather than individuals; see Richard Walton and Robert McKersie (1965). A union leader, for example, can try to incite and yet control the rank and file, but internal politics and group dynamics may make the rank and file's reactions beyond the union leader's control. In this case, the union leader may become committed in the sense envisioned by Schelling (1956) and Crawford (1982). Alternatively, when the parties are groups a dispute may occur because of asymmetric information between them about features of their relationship that are of no concern to the arbitrator, including perhaps the career concerns of individual negotiators. In this case, with regard to the learning in the arbitrator's inference problem the parties are appropriately viewed as having symmetric information, even though asymmetric information on another dimension caused the dispute. These possibilities suggest that it may be a mistake to take too literally models that approximate a firm and a union as single-agent bargainers.

APPENDIX I

This appendix establishes the global optimality of the offers that satisfy the first-order conditions (6) and (7), (17) and (18), (26) and (27), and (31) and (32). An explicit proof is given for the union's offer specified in (6). All the other cases are analogous.

Let the union's expected payoff given in (2) be denoted

$$\text{(A1)} \quad U(y_u|y_e) = y_e F(\bar{y}) + y_u[1-F(\bar{y})],$$

where $\bar{y} = (y_u + y_e)/2$ is the average of the offers and $F(z)$ is the normal distribution with mean M and precision H. In this notation, the first-order condition (4) is simply $U'(y_u|y_e) = 0$, where the prime denotes a derivative with respect to y_u, and

$$\text{(A2)} \quad U'(y_u|y_e) = [1-F(\bar{y})] - \tfrac{1}{2}(y_u - y_e) f(\bar{y}).$$

The goal of this appendix is to prove that $y_u = M + \sqrt{\pi/2H}$ achieves the global maximum of $U(y_u|y_e^*)$, where $y_e^* = M - \sqrt{\pi/2H}$ from (7).

It suffices to prove that U is quasi-concave in y_u when $y_e = y_e^*$. This is done in two parts. The first establishes that U is concave for $y_u \in [y_e^*, M+D]$, where $D > 0$. And the second establishes that $U' < 0$ for $y_u \in (M+d, \infty)$, where $0 < \sqrt{\pi/2H} < d < D$. (Recall that union offers below the employer's offer are suboptimal for the union, and so need not be considered here.)

The concavity of U on $[y_e^*, M+D]$ follows from evaluating U'' on this interval:

$$\text{(A3)} \quad U''(y_u|y_e^*) = -f(\bar{y}^*) - \frac{1}{4}(y_u - y_e^*)\frac{df(\bar{y}^*)}{dz},$$

where $\bar{y}^* = (y_u + y_e^*)/2$. For the normal distribution with mean M and precision H,

$$-\frac{df(\bar{y}^*)}{dz} = H(\bar{y}^* - M) f(\bar{y}^*), \text{ so}$$

$$\text{(A4)} \quad U''(y_u|y_e^*) = f(\bar{y}^*)\left\{-1 + \frac{H}{4}(y_u - y_e^*)(\bar{y}^* - M)\right\},$$

which is negative for

$$(y_u - M)^2 < \frac{1}{H}\left(8 + \frac{\pi}{2}\right) = D^2.$$

The negative slope of U on $(M+d, \infty)$ follows from a simple inequality:

$$\text{(A5)} \quad H(\bar{y}^* - M)[1 - F(\bar{y}^*)] < f(\bar{y}^*).$$

This inequality follows from integrating $(x - \bar{y}^*)f(x)$ over $(\bar{y}^*, \infty)$ when x is normal with mean M and precision H. Given (A2) and (A5), it suffices to show that

$$\text{(A6)} \quad (y_u - y_e^*)(\bar{y}^* - M) > \frac{2}{H},$$

which holds for

$$\text{(A7)} \quad (y_u - M)^2 > \frac{1}{H}\left(4 + \frac{\pi}{2}\right) = d^2.$$

APPENDIX II

This appendix proves Proposition 2. The proof that the arbitrator's strategy is a sequentially rational best response to the parties' strategies is simple. Using (37), the inference rule $M(y_e, y_u, s_a|k)$ given in Proposition 2 can be expressed as the inference rule $s_p(y_e, y_u, s_a)$ used in (36). This inference rule is clearly correct in equilibrium and is well defined off the equilibrium path; given this belief, (36) guarantees sequential rationality.

The proof that the parties' strategies are best responses to each other and to the arbitrator's strategy is slightly more involved. The central point is that out of equilibrium in the conventional-arbitration game, it is as though the arbitrator makes the parties play a final-offer game that has a unique Nash equilibrium that yields both parties the expected payoff $M(s_p|k)$, which equals their equilibrium expected payoff in conventional arbitration. Consider the payoff to, say, the employer from submitting an offer different from the equilibrium offer given by (32). (A deviation by the union is analyzed symmetrically.) By hypothesis, the union's offer will be given by (31), so any deviation by the employer will cause $s_{pe}(y_e|k) \neq s_{pu}(y_u|k)$. This in turn causes the arbitrator to use one of the two out-of-equilibrium parts of the inference rule $M(y_e, y_u, s_a|k)$ specified in the Proposition 2. Therefore, the problem of finding the optimal misleading claim, s_{pe}, for the employer to submit in the conventional-arbitration game is equivalent to the problem of finding the optimal offer, y_e, for the employer to submit in the final-offer game analyzed in the text. But the solution to this second problem is given by (32) and yields a payoff identical to the employer's equilibrium payoff in the conventional-arbitration game. Thus, no deviation is profitable.

REFERENCES

Ashenfelter, Orley and Bloom, David, "Models of Arbitrator Behavior: Theory and Evidence," *American Economic Review*, March 1984, *74*, 111–24.

______, Dow, James and Gallagher, Daniel, "Arbitrator and Negotiation Behavior under an Appelate System," Princeton mimeo., August 1986.

Bloom, David, "Empirical Models of Arbitrator Behavior Under Conventional Arbitration," *Review of Economics and Statistics*, December 1986, *68*, 578–85.

______ **and Cavanagh, Christopher,** (1986a) "An Analysis of the Selection of Arbitrators," *American Economic Review*, June 1986, *76*, 408–22.

______ **and** ______ "An Analysis of Alternative Mechanisms for Selecting Arbitrators," Harvard Institute of Economic Research D.P. No. 1224, April 1986b.

Cormick, Gerald and Patton, Leota, *Environmental Mediation: Defining the Process Through Experience*, Seattle: Office of Environmental Mediation, University of Washington, 1977.

Crawford, Vincent, (1982a) "Compulsory Arbitration, Arbitral Risk, and Negotiated Settlements: A Case Study in Bargaining Under Imperfect Information," *Review of Economic Studies*, January 1982, *49*, 69–82.

______, (1982b) "A Theory of Disagreement in Bargaining," *Econometrica*, May 1982, *50*, 607–37.

______ **and Sobel, Joel,** "Strategic Information Transmission," *Econometrica*, November 1982, *50*, 1431–51.

DeGroot, Morris, *Optimal Statistical Decisions*, New York: McGraw-Hill, 1970.

Farber, Henry, "An Analysis of Final-Offer Arbitration," *Journal of Conflict Resolution*, December 1980, *35*, 683–705.

______ **and Bazerman, Max,** "The General Basis of Arbitrator Behavior: An Empirical Analysis of Conventional and Final-Offer Arbitration," *Econometrica*, November 1986, *54*, 819–44.

Fisher, Roger, *International Mediation: A Working Guide*, New York: International Peace Academy, 1978.

Grossman, Sanford and Hart, Oliver, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, August 1986, *94*, 691–719.

Joskow, Paul, "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets," *American Economic Review*, March 1987, *77*, 168–85.

Kalai, Ehud and Rosenthal, Robert, "Arbitration of Two-Party Disputes Under Ignorance," *International Journal of Game Theory*, 1979, *7*, 65–72.

Kochan, Thomas, "Collective Bargaining and Organizational Research," in Barry Staw and L. Cummings, eds., *Research in Organizational Behavior*, Vol. 2, Greenwich, CT: JAI Press, 1980.

______ **and Jick, Todd,** "The Public Sector Mediation Process: A Theory and Empirical Examination," *Journal of Conflict Resolution*, June 1978, *22*, 209–38.

Nash, John, "The Bargaining Problem," *Econometrica*, January 1950, *18*, 155–62.

Rubinstein, Ariel, "Perfect Equilibrium in a Bargaining Model," *Econometrica*, January 1982, *50*, 97–109.

Schelling, Thomas, "An Essay on Bargaining," *American Economic Review*, June 1956, *46*, 281–306.

Stevens, Carl, "Is Compulsory Arbitration Compatible with Bargaining?," *Industrial Relations*, February 1966, *5*, 38–52.

Walton, Richard and McKersie, Robert, *A Behavioral Theory of Labor Negotiations*, New York: McGraw-Hill, 1965.

Williamson, Oliver, *Markets and Hierarchies*, New York: Free Press, 1975.

Market Incentives for Safe Commercial Airline Operation

By SEVERIN BORENSTEIN AND MARTIN B. ZIMMERMAN*

Airlines are insured against most direct costs of an accident, but they cannot insure against demand loss. Our estimation of deviations from expected demand following accidents finds little or no effect prior to airline deregulation and weak indication of a response to recent crashes. These results are consistent with the changes in an airline's equity value following an accident, which are statistically significant, but quite small relative to the total social cost of the accident.

Central to the debate over the need for government regulation of product safety is the incentive that the private market provide for producing safe goods and services. This incentive comes from the costs imposed upon firms responsible for unsafe products. With the economic deregulation of airlines, the industry has become a focus for this debate. Some argue that increased competition may lead airlines to skimp on investments in safety.[1] Others respond that airline safety is still monitored by the Federal Aviation Administration (FAA) and that the private market discipline unsafe operations, as well. Using a sample of nearly all fatal U.S. airline accidents between 1960 and 1985, this paper attempts to quantify the costs that airlines incur due to crashes. The potential costs include those due to loss of life and equipment, tort liability, increased regulatory oversight, and loss of demand as consumers turn away from products perceived to be unsafe.

We begin by examining the losses incurred by shareholders when a major accident occurs. Airlines carry insurance against many of the costs of a crash, such as equipment loss and tort liability. Hence, the decline in firm value comes from those losses that are uninsured and from the increased insurance costs that might result as insurance companies update their information about the safety of a particular airline or of the industry in general.

A potentially important uninsured loss is the reduction in demand that an airline might suffer. Consumers may respond to an accident in several different ways. If they interpret the news as evidence that flying in general is more dangerous than previously thought, total industry demand may decline. To the extent that all airlines suffer from one carrier's accident, the firms have an interest in enforcing common safety standards. If, however, the information is interpreted as firm-specific, demand for the involved airline may fall while total traffic of competing airlines increases as passengers switch carriers. Finally, consumers may exhibit little or no reaction to an accident, either because they believe that there is no new information in the event, or because the change in their subjective probabilities of an accident has very little effect on the overall value they receive from the product.

We find that an airline's shareholders suffer a statistically significant wealth loss when the airline experiences a serious accident. The average loss in equity value, however, is

---

*Department of Economics and Institute of Public Policy, University of Michigan, Ann Arbor, MI 48109-1220, and School of Business Administration, University of Michigan, Ann Arbor, MI 48109-1234, respectively. For helpful comments, we are indebted to Jeff MacKie-Mason, Michael Bradley, Nancy Rose, participants in seminars at the University of Michigan and Northwestern University, and an anonymous referee. We also thank for excellent research assistance Douglas Couchman, Dan Immergluck, Geoffrey Joyce, Diane Kowalski, Sue Majewski, Vivian Murai, M. Nimalendran, Bruce Schillo, Jon Western, George Wisnewski, and especially, Jamie Seguino. Any remaining errors are our own.

[1] Nancy Rose, 1988, addresses this question directly by examining the relationship between safety record and an airline's financial health.

much smaller than the total social costs of an accident, reflecting the fact that airlines are insured against many of the costs of a crash. This result is also consistent with our finding of little or no effect of an accident on demand. Prior to deregulation, the demand that a carrier faced was virtually unaffected when it experienced an accident. We find some evidence of demand effects in the post-deregulation period, but the statistical significance of this result is weak. Finally, in both the pre- and post-deregulation periods, there appears to be very little evidence of an externality effect, positive or negative, caused by one airline's accident on the demand for other carriers' services.

I. Previous Findings of Market Incentives for Safe Products

Several studies have examined incentives provided by the private market for safety in autos, pharmaceuticals, and air travel. Steven Crafton, George Hoffer, and Robert Reilly (1981), and Reilly and Hoffer (1983) examined the effect of product recalls on the demand for automobiles. Their research suggests that in the month after a product recall the demand for the model type subject to the recall is reduced. Further, they found that similar-sized models of other producers are also adversely affected by the product recall. They did not, however, examine the persistence of the effect and the cost to the automobile manufacturer.

Gregg Jarrell and Sam Peltzman (1985) studied the loss to shareholders from a product recall. Using data for the automobile and pharmaceutical industries, they examined equity value changes associated with the announcement of a product recall. Their analysis suggests that recalls lead to substantial wealth loss for shareholders of the firms involved. They found that the total loss sustained due to recalls far exceeds the readily available estimates of the direct costs of the recall. They attributed this large difference to loss of goodwill, for example, the decline in future expected demand, but they did not measure such losses directly. They also found that the financial losses spill over to competing firms not directly involved in the recall. Jarrell and Peltzman's results relating to the auto industry, however, have been called into question recently by Hoffer, Stephen Pruitt, and Reilly (1988), who found that the results are very sensitive to sample selection and econometric technique.

A number of previous studies have examined the effect of airline accidents on the equity value of airlines and aircraft manufacturers. Andrew Chalk (1986, 1987) examined the effects of commercial air crashes on the value of the firm that manufactured the aircraft. The first paper studied the effect of a single crash, the American Airlines DC-10 accident in Chicago on May 25, 1979. Following the accident, McDonnell-Douglas suffered a loss of about $200 million, which, Chalk concludes, exceeds any reasonable estimate of regulatory or liability costs. The results suggest that the market anticipated a decline in future sales of McDonnell-Douglas aircraft.[2]

Chalk's second paper studied a set of 76 accidents, comparing the change in the manufacturer's equity value following each of 23 crashes that were possibly due to defects in the aircraft ("suspect cases") with the value change after the 53 accidents that were due to other causes. He found significant effects when the manufacturer may be at fault, but no change when the accidents are clearly attributable to other factors. In the 19 suspect cases that involved aircraft still in production, the manufacturer's equity value declined by an average of $21 million. Chalk compared the loss in firm value of these 19 accidents with the four suspect cases that involved aircraft no longer in production. The results are not significant for the latter group, from which Chalk concluded, somewhat tentatively, that the change in firm value in the 19 "in-production" cases is driven in part by expectations of lost future sales of the aircraft.[3]

---

[2] We find that the effect of this accident on American Airlines was much smaller than Chalk's estimate for McDonnell-Douglas and the effect on American was not statistically significant.

[3] A statistical test for the difference in means between the 19 suspect "in-production" cases and the 4 suspect "past production" cases fails to reject the hypothesis that the means are equal.

Two recent papers examine the airline equity value response to crashes. Don Chance and Stephen Ferris (1987) and Mark Mitchell and Michael Maloney (1988) each looked at samples of fatal accidents from the 1960s through the mid-1980s. Each study found statistically significant effects of the crashes on the stock of the airline involved. Chance and Ferris found no statistically significant effect of crashes on other airlines. Mitchell and Maloney distinguished between crashes that are the airline's fault and those that are due to some other cause. For the 31 crashes in their sample that were the airline's fault, they found a statistically significant 2.2 percent decline in equity value (2 post-crash trading days). For 18 crashes that were not the airline's fault, they found a 1.2 percent decline in equity value (2 post-crash trading days) that is not statistically significant.[4] Mitchell and Maloney looked more closely at the impact of accidents on insurance premiums and concluded that changes in insurance rates explain about 34 percent of the loss in equity value. The remainder they attributed to expected loss of consumer demand.

Several of these studies attribute much of the lost firm value to declines in future demand for the firms' products. In this paper, we examine directly the consumer response to airline accidents and attempt to measure the financial loss suffered by shareholders because of the consumer demand response. We find small and statistically insignificant demand changes following an accident and show that little if any loss of airline equity value following accidents can be reliably attributed to consumer response.[5] Further, when the proportional stock value changes are translated into changes in total firm value (dollar value changes), it becomes clear that the losses to airlines are quite small.

II. "Rational" Consumer Responses to Airline Accidents

The reaction that one would expect from a consumer following an accident will depend on her prior distribution of beliefs about airline safety. If the distribution were massed at a single point, no updating would occur. The absence of response in this case would be a result of the consumer's strong beliefs that an airline is safe. If the individual's priors are more diffuse, however, updating from a single crash or a series of accidents is more likely.

Table 1 presents an example of such Bayesian updating where the consumer believes that the airline is probably "safe" (probability of a fatal accident is 1 in 500,000 for each flight, about equal to the average rate in the 1970s), but might be "unsafe" (probability of a fatal accident is 1 in 100,000), and she is able to update her beliefs based on a sample of 200,000 flights (which would be 1 year of operations for a medium-sized major airline).[6] Table 1 demonstrates that weaker prior beliefs about the safety of an airline can lead to significant updating based on accident rates that are

---

[4] From these results, they conclude that crashes only affect equity value when the accident is the airline's fault. However, their reported standard errors indicate that a test for differences in means would fail to reject the hypothesis that the average decline in equity value is the same for fault and non-fault accidents.

[5] Mitchell and Maloney take issue with our conclusions, arguing that one would not expect to see changes in quantity. Rather, they assert, one would see airlines respond to a decline in demand by lowering price, either directly to the consumer or indirectly with increased travel agent commissions, or by increasing quality, for example with greater flight frequency, better-quality food, or more courteous flight attendants. We control for price changes, though they are unlikely to matter much under regulation, because prices were set exogenously by the Civil Aeronautics Board (CAB) and changed infrequently. Likewise, during the regulation period from which most of our sample is drawn, travel agent commissions were set by the CAB. Though flight frequency on a given route may be altered in the short-run, systemwide flight frequency for an airline is constrained by its total fleet size, which cannot be changed substantially in 3 to 4 months without high transaction costs. Though we are unable to control for food quality or employee courtesy, a demand response that could be offset by such changes would not be a significant change to begin with.

[6] At this rate of operations, a "safe" carrier would average one fatal accident every 2 1/2 years and an "unsafe" airline would average 2 fatal accidents per year.

TABLE 1—EXAMPLE OF BAYESIAN UPDATING FROM ACCIDENT RECORD

Number of Accidents in One Year	Prior Beliefs		
	Pr(safe)[a] = 0.9 Pr(Unsafe)[b] = 0.1	Pr(Safe) = 0.99 Pr(Unsafe) = 0.01	Pr(Safe) = 0.999 Pr(Unsafe) = 0.001
	Posterior Beliefs		
0	Pr(Safe) = 0.978 Pr(Unsafe) = 0.022	Pr(Safe) = 0.998 Pr(Unsafe) = 0.002	Pr(Safe) = 1.000 Pr(Unsafe) = 0.000
1	Pr(Safe) = 0.899 Pr(Unsafe) = 0.101	Pr(Safe) = 0.990 Pr(Unsafe) = 0.010	Pr(Safe) = 0.999 Pr(Unsafe) = 0.001
2	Pr(Safe) = 0.641 Pr(Unsafe) = 0.359	Pr(Safe) = 0.951 Pr(Unsafe) = 0.049	Pr(Safe) = 0.995 Pr(Unsafe) = 0.005
3	Pr(Safe) = 0.263 Pr(Unsafe) = 0.737	Pr(Safe) = 0.797 Pr(Unsafe) = 0.203	Pr(Safe) = 0.975 Pr(Unsafe) = 0.025
4	Pr(Safe) = 0.067 Pr(Unsafe) = 0.933	Pr(Safe) = 0.440 Pr(Unsafe) = 0.560	Pr(Safe) = 0.888 Pr(Unsafe) = 0.112

Note: Updating based on accident record in 1 year, assumed = 200,000 flights.
[a]"Safe" = A 1 in 500,000 probability of a fatal accident on any flight.
[b]"Unsafe" = A 1 in 100,000 probability of a fatal accident on any flight.

actually observed for some airlines. Furthermore, the likelihood of significant changes in one's beliefs is much greater if multiple accidents are observed in a sample period.

The small consumer response that we have estimated could be a result of strong prior beliefs about the safety of air travel or confidence that the FAA will react to an accident with much closer monitoring. Alternatively, the weak reaction is also consistent with a significant change in subjective probabilities, but a low marginal valuation of safety. Because flying is an extremely safe form of travel, increases in the perceived danger of flying may have little effect on a consumer's expected cost of travel. The decrease in expected consumer surplus due to a substantial increase in subjective accident probabilities may still leave most of an airline's customers with positive consumer surplus from the product. The data do not allow us to distinguish among these possible explanations for weak demand responses to accidents.

III. Method of Study

The first part of this research examines the effect of an airline accident on the equity value of the airline. Drawing on the efficient markets literature (Eugene Fama, Lawrence Fisher, Michael Jensen, and Richard Roll, 1969), the change in equity value associated with an airline accident is taken as an unbiased estimate of the financial consequences of the accident. The market model is used to separate out changes in value caused by overall market effects from those changes caused by the accident itself.

The normal relation between the returns to a given stock and the market is given by

$$R_{jt} = \alpha_j + \beta_j R_{mt} + \varepsilon_{jt}. \tag{1}$$

The parameter β_j measures the sensitivity of the jth firm's return, R_{jt}, to movements in the market index, R_{mt}, and is equal to β of the Sharpe-Litner capital asset pricing model. The term $\beta_j R_{mt}$ is the portion of the return to security j on day t that is due to marketwide factors. The parameter α_j measures that part of the average daily return on the stock that is not due to market movements. Lastly, ε_{jt} measures that part of the change in the value of firm j's stock on day

t that is not due to either movements in the market or to the firm's average daily return.[7]

On the day of an event, the deviation in an individual stock's daily return from what is expected based on equation (1), that is, the prediction error, is taken as an unbiased estimate of the financial effects of the event. Let PE stand for this prediction error:

$$(1') \qquad PE_{jt} = R_{jt} - a_j - b_j R_{mt},$$

where a_j and b_j are, respectively, the estimates of α_j and β_j.

The average abnormal daily return for all accidents in the sample is calculated along with two measures of its statistical significance.[8] The first measure of significance aggregates into a single portfolio the abnormal returns of all airlines experiencing an accident for the day of each firm's crash. It then uses the daily variance of returns on this portfolio to calculate a t-statistic. Assuming that the N abnormal returns are independently distributed, the formula for this t-statistic is given by

$$(2) \qquad t = \frac{\Sigma_1^N PE_{jt}}{\sqrt{\Sigma_1^N VAR(PE_{jt})}},$$

where $VAR(PE_{jt})$ is the variance of the prediction error for firm j taken from the market model regression of equation (1).

The second significance test calculates a t-statistic for each firm's abnormal return for each accident-day. The sum of these individual t-statistics follows a distribution that is asymptotically normal with mean zero and variance equal to the number of observations. The z-statistic for the average is then the sum of the individual t-statistics divided by the square root of the number of observations. This test attributes less weight to observations of firms with a high variance in returns and is therefore less sensitive to distortions from very noisy observations. The formula for this z-statistic is

$$(3) \qquad z = \frac{\Sigma_1^N \left(PE_{jt} / \sqrt{VAR(PE_{jt})} \right)}{\sqrt{N}}.$$

The share of the financial losses that can be attributed to demand response is examined in the following manner. For every airline for which sufficient monthly traffic data can be obtained, a demand function is estimated during months unaffected by the airline's accidents. Demand is modeled as a function of price, income, a time trend, and seasonal (dummy) variables. The month in which an airline experiences an accident and the three following months are excluded from the estimation of these parameters. The deviation of actual demand from predicted demand in these four monthly periods is the measure of demand response to the accident, and is estimated directly using dummy variables for each of the months. The function estimated for each firm during each time period is

$$(4) \qquad \ln RPM_{jt} = \lambda_j + \phi t + \alpha_j \ln P_{jt} + \beta_j \ln I_t + \sum_{i=1}^{11} \gamma_{ij} F_{it} + \sum_{l=1}^{L_j} \sum_{k=0}^{3} \delta_{jlk} C_{jlk} + \varepsilon_{jt}$$

where:

RPM_{jt} = Revenue passenger miles for firm j in month t.

---

[7] The market model regressions were estimated using the 280 trading days before the accident. If 280 days of trading data before the date of the accident were not available, the model was estimated using the n days of data available before the accident and $280-n$ days beginning 40 days after the accident. The market return variable used was the equally weighted index. In a number of cases, the 280 days for one firm's market regression included an accident by another firm. The bias from such an overlap is almost surely negligible, however, because even the own-firm effects are found to last only one or two days, and because Chance and Ferris find the cross-airline equity value effects to be insignificant. See Section V, Part E.

[8] The day of the crash is counted as the first day of trading on the information if the crash occurred before 3 P.M. New York City local time and it was a trading day. Otherwise, the first day of the effect was taken to be the next trading day.

t = The number of the month within either the 72- or the 96-month sample period (see Section IV).
P_{jt} = Average revenue per passenger-mile (yield) for firm j in month t.
I_t = U.S. Personal income in month t.
F_{it} = Eleven seasonal dummy variables, each of which takes on a value of one in month i and zero otherwise, February omitted.
C_{jlk} = A vector that takes on the value 1 in the kth month after the lth crash that firm j experienced during the time period and zero otherwise. ($k = 0$ refers to the month in which the crash occurred.)
λ_j = The estimated constant term for firm j.
ϕ_j = The estimated time trend for firm j.
α_j = The estimated price elasticity of demand faced by firm j.
β_j = The estimated income elasticity of demand faced by firm j.
γ_{ji} = The estimated natural log of the proportional deviation of demand from the base month (February) in month i for firm j.
δ_{jlk} = The natural log of the proportional prediction error of the regression for demand k months after crash l experienced by firm j during the time period.

Equation (4) is estimated for each airline separately.[9] The equation is estimated for each of four time periods using a Cochrane-Orcutt procedure to correct for serial correlation in the disturbances. For the three time periods during which the airline industry was regulated by the CAB—1960–65, 66–71, and 72–77—price is treated as an exogenous variable. During these years, price was determined by regulators and it is assumed that price did not respond to short-run demand changes. Beginning around 1978, however, airlines were given much greater pricing freedom that ultimately led to complete deregulation. For this latter period, 1978–85, the demand equation is also estimated by two-stage least squares (2SLS). The instruments for price include the exogenous variables listed above and an index of airline costs that is described in the Data Appendix.

Analogous to the significance test of stock movements, we use two different approaches to test the significance of the impact of accidents on demand in the month of the crash and the three succeeding months. In each of the four post-crash months, we calculate the average of the estimated log proportional impact of a crash from the estimates of the δ parameters.[10] We also calculate the standard error of this estimate of the mean and perform a t-test of whether the log deviations from predicted demand in the months of the crashes and the first, second, and third following months are statistically different from zero.

As with the financial market study, such t-tests of the mean can be very sensitive to a few high-variance estimates. Again, this leads to the alternative calculation of a z-statistic for a test of statistical significance. For the "portfolio" of accidents, the following z-statistic is calculated for the month of the crash and each of the three succeeding months:

---

[9] Though the structure of the model appears to allow for use of Zellner's seemingly unrelated regressions technique, its use in this case would distort estimation of the effect of a crash. Unless one is willing to assume away externality effects on demand, correcting the estimated parameters for firm j, which experiences a crash at time t, based on the residuals in the estimates of other firms' demand functions at time t will inappropriately dampen (or enhance, if the externality effect is positive) the estimated effect of the crash on the demand faced by firm j. Zellner's *SUR* could be used if one excluded from all firms' regressions the periods following any firm's accident. This would eliminate too many of the observations to make demand estimation possible.

[10] In the two-stage least squares demand estimation for the deregulation period, there is a technical issue of whether one should use the residuals from the second-stage estimation (with $\hat{p}$ as a right-hand side variable) or, having estimated the parameters of the demand function, one should calculate the residuals from the expected quantity using the actual values of the endogenous variable, p. The latter approach would yield estimates of the *proportional change in the quantity demanded* due to the accident, while the former approach gives the *proportional shift in the demand function*. We use the former approach and thus measure the proportional change in quantity *for a given price*.

$$(5)\quad z_k = \frac{\sum_{j=1}^{J}\sum_{l=1}^{L_j}\left(\delta_{jlk}/\sqrt{VAR(\delta_{jlk})}\right)}{\sqrt{\sum_{j=1}^{J}L_j}},$$

where j indexes firms, l indexes crashes, k indexes the month following a crash, J is the number of airlines experiencing at least one crash during the period, and L_j is the number of crashes experienced by airline j. These tests attribute less weight to estimated demand effects from airlines with more volatile demands.

For each accident, the four monthly deviation parameters, δ_0 through δ_3, can be translated into percentage deviations from predicted demand. The sum of these percentage deviations indicates the magnitude of the total lost demand as a percentage of one month's traffic. This interpretation is not exact, because demand fluctuates seasonally and cyclically, so the percentage deviations in each of the four months are not deviations from the same predicted level. Still, the sum gives an indication of the overall demand effect of an accident.

Following estimation of the demand responses, an attempt is made to explain the variation in responses. The estimated δ_{jlk} coefficients are regressed on a set of variables explaining the magnitude of the demand effect during each month k after the crash. The explanatory variables are the number of fatalities (*FATAL*), the size of the airline as measured by revenue passenger-miles flown in the month before the crash (*RPMLAG*), whether the airline was primarily responsible for the accident (*FAULT*), the day of the month on which the crash occurred (*DATE*),[11] the recent accident records of the firm (*FATSUMOWN*), and the industry (*FATSUM*),[12] and two measures of the extent of newspaper coverage of the crash. These measures are the number of days of front page coverage (*PAGE1*) and the total number of articles appearing in the *New York Times* in the two weeks after the crash (*ART*).[13]

The regression of the estimated demand effects, δ_{jlk}, on causal variables allows calculation of a fitted value for demand change. The fitted value, $\hat{\delta}_{jlk}$, along with the quantity that is actually sold, can be used to calculate the quantity that the model would have predicted in the absence of the accident. The difference between the predicted and the actual quantity is an estimate of the loss in passenger traffic due to the crash.[14] This loss multiplied by the average yield (price) in the period is an estimate of revenue loss attributable to the accident.[15] The estimated revenue loss is then used as an independent variable in regressions that explain the change in the airlines' equity value. The *fitted* value for the demand response (*REVCHFIT*) is used in order to purge the estimate of errors due to nonsystematic influences on demand experienced in the months after a crash.[16]

Because the dependent variable in this procedure is the stock movement for the first and second trading days following the accident, estimates by traders of the likely demand impact of a crash can only be based on information available at that time. Thus, the extent of newspaper coverage in the following two weeks and whether or not the accident was primarily the carrier's fault are excluded as instruments in forming *REVCHFIT*.[17]

---

[11] If the demand impact decays monotonically over time, one would expect this variable to be positively correlated with the month-of-crash impact and negatively correlated with the impact during all later months. These signs may seem counterintuitive because a larger crash effect is a more negative abnormal demand.

[12] More precisely, these are the number of fatalities for, respectively, the firm and the industry in the 365 days prior to the accident.

[13] Because crashes in New York are covered more extensively by the *Times* than crashes elsewhere, newspaper coverage variables for New York crashes are taken from the coverage of the *Chicago Tribune*.

[14] $RPM - R\hat{P}M = RPM\cdot(\hat{\theta}_{jlk}/(1+\hat{\theta}_{jlk}))$, where $\hat{\theta}_{jlk}$ is the estimated percentage change in demand, $\hat{\theta}_{jlk} = \exp\hat{\delta}_{jlk} - 1$.

[15] In the deregulated period, the traffic loss is multiplied by the fitted price, $\hat{p}$, from the first stage of the two-stage least squares demand estimation procedure. Due to the brief time period over which the demand response seems to persist, when it is evident at all, we ignore the effect of discounting in analyzing the impact of lost revenue on the present value of future profits.

[16] Essentially, this is an instrumental variables procedure to correct for errors in a right-hand side variable.

The variables used to estimate the $\hat{\delta}_{jlk}$ explain the variation in demand response rather poorly. Hence, revenue changes calculated from the actual estimated demand deviations (*REVCHACT*) are also used to explain the change in firm value. The results do not differ substantially for the *REVCHACT* and *REVCHFIT* variables. To obtain the dependent variable in these regressions, the percentage financial losses, as measured by the market model residuals, are converted to an absolute dollar loss. These losses are then regressed on the estimated loss in revenue and on other factors reflecting the direct costs of an accident.

The last issue addressed is the externality effect, the impact of one airline's accident on the demand faced by other airlines and on the equity values of other airlines. The approach is quite similar to the own-firm analysis. The abnormal return on equity for all major airlines other than the one experiencing the accident is estimated for the accident day and the following day. Similarly, the deviation from predicted demand for all other airlines in the month of the accident and the following month is calculated.[18] The residual from a carrier's demand equation in the month of another airline's crash is the best estimate of the impact of another carrier's accident on the quantity sold by the first airline.[19] Test statistics for demand and financial effects, analogous to the test statistics for own-firm effects described above, are then calculated.

IV. Description of the Sample

The base sample from which all the analysis is drawn is every accident aboard a U.S. certificated air carrier[20] from 1960 to 1985 involving at least one on-board fatality and some damage to the aircraft.[21] Demand or stock market data were not available for every firm at all times during the 26-year period, however, so the demand and financial studies omit some of these accidents. Table 2 lists all accidents included in either the demand or the financial analysis.

Only accidents occurring between 1962 and 1985 are included in the estimation of financial losses associated with airline accidents. This time period is chosen in order to allow use of the daily stock returns tape of the Center for Research in Security Prices (CRSP) of the University of Chicago. Use of this data source also limits the investigation to firms listed on either the New York or American Stock Exchanges. Seventy-four accidents met these criteria, but 7 were eliminated from the regression of value loss on injuries and revenue loss, because there was insufficient demand data to allow estimation of the effect of the crash on demand.[22]

---

[17] In place of the newspaper variables described above, a dummy variable for page 1 coverage on the day following the accident was used. Although the cause of the accident is occasionally known by investigators within the 48 hours following the accident, it is rarely stated publicly with any degree of certainty during this time.

[18] Because we found very little evidence of a demand externality in these periods, we did not extend the analysis for later months.

[19] In the demand analysis, months in which other airlines experienced accidents are not excluded from the estimation of an airline's demand functions. Because these observations are included, problems would arise if there were a significant demand externality and/or if other airlines' crashes were correlated with variables in the demand regressions. The combination of these two factors would create omitted variable bias in the parameter estimates for the included variables. In fact, accidents are not significantly correlated with any included variables. Even without such a correlation, omitting consideration of a demand externality, if it were present, would bias the reported standard errors in the demand regressions. As discussed in Section V, Part E, we find very little evidence of a demand externality.

[20] These are Part 121 certificated airlines and thus exclude air taxis and commuter operations.

[21] The dual conditions of on-board fatalities and aircraft damage eliminates from the sample those fatalities not directly related to air travel, for example, heart attacks, and fatalities of ground crew or other bystanders.

[22] In two of the seven cases eliminated, the same carrier had two crashes during the same month, so we

TABLE 2—LISTING OF ACCIDENTS AND INCLUSION IN ANALYSIS

Date	Carrier	Aircraft	Damage	FTL	SER	TOT	*VALCH12* ($million)	Notes
600106	NA	DC-6B	D	34	0	34		X
600317	NW	L-188C	D	63	0	63		X
600523	DL	Convair 880	D	4	0	4		EX
600722	TX	DC-3	S	1	0	12		X
601004	EA	L-188	D	62	9	72		X
601028	NW	DC-4	D	12	0	12		AX
601216	TW	L-1049	D	44	0	44		AX
601216	UA	DC-8	D	84	0	84		X
610128	AA	B-707	D	6	0	6		AEX
610711	UA	DC-8	D	17	12	122		X
610721	AS	DC-6	D	6	0	6		DEX
610901	TW	L-049	D	78	0	78		X
610917	NW	L-188C	D	37	0	37		X
620301	AA	B-707	D	95	0	95		X
620522	CO	B-707	D	45	0	45		X
621019	AL	Convair 440	M	1	0	52	0.913	
621123	UA	Viscount	D	17	0	17	3.603	
621130	EA	DC-7	D	25	14	51	7.049	
630129	CO	Vickers 810	D	8	0	8	−0.083	
630212	NW	B-720B	D	43	0	43	9.503	
630603	NW	DC-7	D	101	0	101	−8.317	N
630702	MK	Martin 404	D	7	30	43	−0.224	
631208	PA	B-707	D	81	0	81	−2.575	
640225	EA	DC-8	D	58	0	58	13.842	
640312	FL	DC-3C	D	5	0	5		X
640709	UA	Viscount 745D	D	39	0	39	−40.560	
641115	BZ	Fairchild F-2	D	29	0	27		X
641123	TW	B-707 331	D	48	11	73	−5.003	
650208	EA	DC-7B	D	84	0	84	36.956	
650816	UA	B-727	D	30	0	30	−26.997	
650917	PA	B-707 121B	D	30	0	30	−6.324	
651108	AA	B-727	D	58	4	62	−18.300	
651111	UA	B-727	D	43	35	91	−107.327	
651204	EA	L1049C	D	4	34	54	−8.668	
660806	BN	British AC	D	42	0	42	−0.262	
661115	PA	B-727	D	3	0	3	52.299	DE
661120	PI	Martin M404	D	3	0	3		DEX
670309	TW	DC-9	D	25	0	25	39.824	
670330	DL	DC-8	D	6	0	6	−51.000	E
670623	MK	British AC-111	D	34	0	34	−8.225	
670719	PI	B-727	D	79	0	79		X
671106	TW	B-707	D	1	1	36	−20.959	O
671121	TW	Convair 880	D	69	13	82	−9.459	
671221	FL	DC-3C	D	2	0	2	−2.302	DE
680503	BN	L-188	D	85	0	85	54.942	
680612	PA	B-707	D	6	0	62	−32.354	
680810	PI	Fairchild 227	D	35	2	37		X
681025	NE	Fairchild 227	D	32	7	49	−19.973	
681212	PA	B-707	D	51	0	51	−107.342	
681224	AL	Convair 580	D	20	26	47	−7.372	
681226	PA	B-707	D	3	0	3	137.813	ODE
681227	NC	Convair 580	D	27	16	45		X
690106	AL	Convair 440	D	11	14	28	−0.899	
690118	UA	B-727	D	38	0	38	19.255	
690726	TW	B-707	D	5	0	5	86.667	E
690909	AL	DC-9	D	82	0	82	5.114	
691119	MK	Fairchild 227B	D	14	0	14	−1.311	
701114	SO	DC-9	D	75	0	75		X
710331	WA	B-720	D	5	0	5	−8.751	E

TABLE 2—CONTINUED

Date	Carrier	Aircraft	Damage	FTL	SER	TOT	*VALCH12* ($million)	Notes
710607	AL	Convair 580	D	28	3	31	−2.833	
710725	PA	B-707	D	4	0	4	17.336	DE
710904	AS	B-727	D	111	0	111	−5.490	
720303	MK	Fairchild 227B	D	16	32	48	0.559	A
720530	DL	DC-9	D	4	0	4	29.440	E
720629	NC	Convair 580	D	8	0	8		X
721208	UA	B-737	D	43	12	60	20.356	
721220	NC	DC-9	D	10	9	45		X
721229	EA	L1011	D	99	60	176	−168.905	
730722	PA	B-707	D	78	1	79	9.175	
730723	OZ	Fairchild 227B	D	38	6	44	−5.986	A
730731	DL	DC-9	D	88	1	89	−88.244	
730927	TX	Convair 600	D	11	0	11		X
731103	PA	B-707	D	3	0	3	3.634	DE
731103	NA	DC-10	S	1	0	128	−16.090	
740131	PA	B-707	D	96	5	101	−17.891	
740422	PA	B-707	D	107	0	107	2.798	
740908	TW	B-707	D	88	0	88	−8.877	
740911	EA	DC9-31	D	71	10	82	−19.389	
741201	TW	B-727	D	92	0	92	3.794	
741201	NW	B-727	D	3	0	3	18.236	E
750624	EA	B-727	D	112	12	124	0.933	
760405	AS	B-727	D	1	11	50	−2.927	
760427	AA	B-727	D	37	19	88	15.719	
770327	PA	B-747	D	327	55	396	−7.481	
770404	SO	DC-9	D	62	22	85		X
771218	UA	DC-8F	D	3	0	3	2.309	E
780118	FL	DHC-6	D	3	0	3	−0.535	E
780301	CO	DC-10	D	2	31	197	−2.260	
780508	NA	B-727	S	3	11	58	−2.154	A
780925	PS	B-727	D	135	0	135	−3.270	A
781228	UA	DC-8	D	10	23	198	−46.282	
790212	AL	Frakes M298	D	2	8	25	1.206	
790525	AA	DC-10	D	271	0	271	−7.714	
791031	WA	DC-10	D	72	13	89	−1.065	
820113	AF	737-222	D	74	5	79		X
820123	WO	DC-10	D	2	4	345	−1.785	
820709	PA	B-737-235	D	145	0	145	−14.234	
820811	PA	B-747	S	1	0	288	−1.146	O
830109	RC	Convair 580	S	1	1	33	−7.410	
830111	UA	DC-8-54F	D	3	0	3	27.214	E
850802	DL	L-1011-385-1	D	134	15	164	−27.142	

Notes. X = Insufficient equity price data to estimate equity value change. A = Insufficient demand data to estimate demand effect. O = Excluded from demand effect estimation due to complete overlap of observation with another accident by same airline.

D = Cargo, E = crew only, N = military charter.

Damage Notes: D = Destroyed, S = Substantial, M = Minor.

FTL = Fatalities; SER = Seriously injured; TOT = Total on board; *VALCH12* is in January 1985 constant dollars.

TABLE 3—CARRIERS INCLUDED IN THE DEMAND AND/OR FINANCIAL STUDIES

AA—American
AF—Air Florida
AL—US Air (formerly Allegheny)
AS—Alaska Air
BN—Braniff
BZ—Bonanza
CO—Continental
DL—Delta
EA—Eastern
FL—Frontier
MK—Mohawk
NA—National
NC—North Central
NE—Northest
NW—Northwest Orient
OZ—Ozark
PA—Pan Am
PI—Piedmont
PS—Pacific Southwest Airlines
RC—Republic
RW—Hughes Air West
SO—Southern
TW—TWA
TX—Texas International (formerly Trans-Texas)
UA—United
WA—Western
WO—World

Estimation of the demand functions was broken into four periods, 1960–65, 1966–71, 1972–77, and 1978–85. The first period covers the time of transition from piston engine to jet aircraft. The next period was relatively stable in both prices and technology. The third period covers the time of the CAB's Domestic Passenger Fare Investigation (DPFI) as well as increased use of wide-bodied aircraft.[23] The final period is taken to be the deregulation era, though the exact starting date can, of course, be questioned.[24] Although there are 27 airlines included in the demand study, every carrier is not included in every period. For an airline to be included during a period, data for at least 48 months of full operation during the period had to be available.[25] In the four periods, there are respectively, 15, 14, 12, and 12 airlines that experienced accidents and for which sufficient data are available. Table 3 lists the airline carriers included in the demand and/or financial studies.

V. Results of the Statistical Analysis

A. *The Average Effect of Accidents on Shareholder Wealth*

Table 4 presents the average abnormal returns for the portfolio of 74 accidents included in the financial analysis from two days before news of the accident reaches the market to 14 days after. The column labeled "CAR" shows the cumulative abnormal returns for the accident day and the following trading days. The *t*- and *z*-tests are presented for the daily and cumulative abnormal returns.[26]

Crashes are associated with, on average, a 0.94 percent loss in the equity value of the firm on the first trading day. This loss is statistically significant at the 1 percent level.[27] The percentage of firms experiencing

---

eliminated the smaller crash from the sample. In both cases, the accident eliminated resulted in only one fatality. In all but one case of less than complete overlap between two accidents by the same airline, the fact of the overlap was ignored. Essentially, this procedure overestimates the demand impact of accidents by attributing the joint effect of two crashes to each of the accidents. In the one such case that exists in the deregulation period, two Pan Am crashes in July and August 1982, the latter accident, involving a bomb explosion that killed one person, was eliminated from the demand studies. Because we found no significant demand effect in the regulation period, we did not explore further corrections for the 4 overlap cases that occurred during this period.

[23] The DPFI was a wide-ranging investigation into airline prices that ran from 1971 to 1974. It resulted in the imposition of new and stricter policies on discount fares and route entry.

[24] The United Airlines crash on 12/18/77 is included in the deregulation period.

[25] Months in which a carrier experienced a labor strike were not considered full-operation periods.

[26] Because there is little indication of an equity value effect after the first day and because the event window can be so confidently specified, we have not done corrections for the interdependence of daily abnormal return estimates that may bias cumulative abnormal return estimates over long event windows. See Michael Salinger, 1988.

[27] This proportional loss is in line with the first-day effect that Chance and Ferris found, −1.18 percent, and that Mitchell and Maloney, 1988, found −1.19 percent. The differences are explained mostly by differences in the sample. Ours extends back to include earlier accidents that either of the other works, but not forward past 1985. Eighteen accidents in our sample do

TABLE 4—FINANCIAL RETURNS ASSOCIATED WITH ACCIDENTS

Day	AR	CAR	Percent Negative	*T*	*T*-CAR	*Z*	*Z*-CAR
−2	−.00009		52.703	−0.28		−.708	
−1	.00104		52.703	.314		−0.22	
0	−.00940	−.00940	64.865	−2.818	−2.818	−3.087	−3.087
1	−.00031	−.00971	52.703	−.093	−2.059	−.108	−2.259
2	.00098	−.00873	51.351	.295	−1.511	.199	−1.730
3	−.00102	−.00975	56.757	−.309	−1.463	−.024	−1.510
4	−.00302	−.01277	62.162	−.910	−1.715	−1.086	−1.837
5	.00090	−.01187	48.649	.271	−1.455	.069	−1.648
6	.00089	−.01098	54.054	.267	−1.246	.434	−1.362
7	.00348	−.00750	50.000	1.050	−.795	.756	−1.007
8	.00075	−.00675	54.054	.227	−.674	−.244	−1.031
9	−.00100	−.00775	52.703	−.302	−.735	−.300	−1.073
10	.00541	−.00234	45.946	1.618	−.213	1.846	−.466
11	−.00240	−.00474	55.405	−.716	−.410	−.874	−.699
12	−.00031	−.00505	55.405	−.092	−.420	.016	−.667
13	.00075	−.00429	48.649	.226	−.344	.544	−.497
14	−.00170	−.00600	54.054	−.510	−.464	−.513	−.613

Notes: AR = Abnormal return; CAR = Cumulative abnormal return; %NEG = percent of firms with negative abnormal returns; *T* = *t*-statistic as described in text; *T*-CAR = The *t*-statistic for the significance test on the cumulative abnormal returns; *Z* = *z*-statistic as described in text; *Z*-CAR = The *z*-statistic for the significance test on the cumulative abnormal returns.

negative abnormal returns is high, 65 percent, which is statistically different from 50 percent at the 5 percent level. It also appears that the information is, on average, totally absorbed in the stock price on the first trading day after the accident. Abnormal returns are small and mostly insignificant after that date. Of course, new information on particular crashes often becomes available as the National Transportation Safety Board conducts its investigation. The average result means that the market forms an unbiased estimate of significant negative consequences as soon as the crash becomes known.

Analysis in terms of the proportional change in equity value could be misleading, however, since debt/equity ratios vary substantially across airlines and many firms have large holdings in non-airline industries. If most of the losses are direct costs from the accident, such as insurance deductibles or co-payments, then the absolute loss of firm value is likely to be more systematically related to crashes than the proportional loss. The average value loss on the first day information about a crash reaches the market is \$3.67 million.[28] The estimated standard error of this mean is \$3.60 million, producing a *t*-statistic of −1.03.[29] The cumulative two-day loss averages \$4.50 million, with a standard error of \$4.49 million and a *t*-statistic of −1.01. The range over these two days is much larger, however, going from a loss of \$169 million to a gain of \$138 million. Clearly, in some cases news about the crash

---

not appear in Mitchell and Maloney's, while 4 in their sample are not included in ours. The sample used by Chance and Ferris is a subset of ours, but omits 25 accidents that we include.

[28] This is in 1985 constant dollars, as are all monetary figures in this paper. These figures include only changes in the value of common stock. If the value of either preferred stock or bonds displayed systematic abnormal returns following accidents, these figures could underestimate the total change in firm value. A large bias seems unlikely, however, because very few of these accidents posed any real threat to the long-term survival of the airline.

[29] The *z*-statistic that results when each value change on the day of an accident is adjusted for the standard error in value change is the same as the *z*-statistic for the proportional change in Table 4, −3.087.

coincides with other new information that has a positive effect on the valuation of the firm.

Of the 74 accidents for which equity value data were available, 12 were crew-only accidents. These are accidents of major carriers, for example, Pan Am and United, but it is possible that they attract less attention from potential passengers and from the FAA, if that agency is guided to some extent by public and congressional pressure. Eliminating these from the sample increases the averages substantially, as the average equity value effect of the 12 crashes removed is positive, though not significant. For the 62 accidents that had passengers on board, the average equity value loss on the first day is −1.29 percent (*t*-stat: −3.58, *z*-stat: −4.01), or −$6.58 million (*t*-stat: −2.03). The average cumulative two-day loss is −1.48 percent (*t*-stat: −2.97, *z*-stat: −3.41), or −$9.97 million (*t*-stat: −2.30). These average equity losses, calculated with or without the crew-only accidents, are small relative to the social losses of the airline accidents. The average death toll in these accidents is more than 40 and in most cases the aircraft destroyed was worth more than $10 million (1985 constant dollars).

B. *The Average Effect of Accidents on Demand*

The results of the four sets of demand regressions are summarized in Table 5. Since 15 to 20 demand regressions were run for each time period (one for each carrier with enough data for estimation within the period) and each included more than 15 right-hand side variables, the results have been summarized in order to make the presentation manageable. The constant terms, trend terms, and seasonal dummy variables are not shown. The firm constants vary as one would expect; larger firms have larger constants. The trend terms vary by airline, but most are positive, particularly in the early periods. As expected, since the omitted month in the seasonal dummy variables is February, the trough in industry demand, most of the carriers show greater demand in every other month and the highest in the summer months. Some north–south carriers, for example, Eastern and National, consistently show a winter peak and lower demand in the summer months.

Though the price and income parameters differ for each carrier, only the weighted averages of these estimates are presented (where the weights are the inverse of the variance of each parameter estimate) along with their standard errors.[30] In the pre-deregulation periods, nearly all price elasticity parameters are negative and the few positive estimates are insignificant. The average elasticities correspond roughly with previous estimates, though they indicate less price-elastic demand than earlier work.[31] The estimated pre-deregulation income elasticity parameters are consistent with previous studies.

Using an industry cost-index as an instrument to identify the effect of price on quantity in the deregulation period, demand functions were estimated for 19 firms. Though the weighted average price elasticity is −0.50, it was positive in 6 estimates, none statistically significant. The weighted average price elasticity estimate when the deregulation period demand functions are estimated by OLS is −0.62, with only one positive (and quite insignificant) estimate. To further check the sensitivity of the results to the price elasticity parameter, the demand functions were re-estimated restricting all firms' price elasticities to be alternatively −0.5 and −2. The results are changed only slightly, as discussed in footnote 35 below.

The primary focus of Table 5 is the deviation from expected demand during the months following a crash. Due to the variation among firms in the precision of the estimates, particularly in the two-stage least squares estimates of the deregulation period, we present both the unweighted average ef-

---

[30] This weighted average approach is identical to estimating the mean by running OLS on a constant and adjusting the regression for heteroskedasticity.

[31] See Richard Ippolito, 1981; Severin Borenstein, 1983; David Graham, Daniel Kaplan, and David Sibley, 1983; and Steven Morrison and Clifford Winston, 1986.

Table 5—Estimated Impact of Accidents on Airline's Demand

1960–65: 31 Observations	1st month	2nd month	3rd month	4th month	Total
Weighted Mean (in Percents)	−0.9	−1.1	−1.0	−1.3	−4.3
Mean	−0.6	0.4	3.0	7.0	9.8
(*t*-statistic)	(−0.48)	(−0.07)	(0.68)	(1.57)	(0.68)
[*z*-statistic]	[−0.56]	[−0.44]	[−0.08]	[0.35]	[−0.41]

Weighted Average Elasticities (Standard Error)—Price: −0.16 (0.04)—Income: 1.09 (0.23)

1966–71: 26 Observations	1st month	2nd month	3rd month	4th month	Total
Weighted Mean (in Percents)	−0.1	−2.5	0.0	1.3	−1.4
Mean (in Percents)	−1.5	−3.1	0.3	1.0	−3.5
(*t*-statistic)	(−1.01)	(−1.60)	(0.03)	(0.35)	(−0.85)
[*z*-statistic]	[−0.52]	[−1.70]	[0.03]	[0.81]	[−0.72]

Weighted Average Elasticities (Standard Error)—Price: −0.59 (0.07) —Income: 1.96 (0.26)

1972–77: 21 Observations	1st month	2nd month	3rd month	4th month	Total
Weighted Mean (in Percents)	0.1	0.6	1.1	−0.2	1.7
Mean (in Percents)	−0.3	−0.2	−.04	−1.3	−2.1
(*t*-statistic)	(−0.33)	(−0.35)	(−0.47)	(−0.85)	(−0.75)
[*z*-statistic]	[0.00]	[0.11]	[0.34]	[−0.42]	[−0.36]

Weighted Average Elasticities (Standard Error)—Price: −0.29 (0.05)—Income: 0.70 (0.11)

1960–77: 78 Observations	1st month	2nd month	3rd month	4th month	Total
Weighted Mean (in Percents)	−0.4	−1.0	−0.1	−0.2	−1.8
Mean (in Percents)	−0.8	−0.9	1.1	2.8	2.2
(*t*-statistic)	(−1.00)	(−1.00)	(0.39)	(1.15)	(−0.14)
[*z*-statistic]	[−0.65]	[−1.20]	[0.14]	[0.47]	[−0.86]

1978–85: 13 Observations (OLS)	1st Month	2nd month	3rd month	4th month	Total
Weighted mean (in Percents)	−1.4	−5.2	−4.7	−3.9	−15.3
Mean (in Percents)	−1.6	−3.1	−1.6	−1.0	−7.4
(*t*-statistic)	(−0.56)	(−0.86)	(−0.45)	(−0.35)	(−0.86)
[*z*-statistic]	[−0.80]	[−1.63]	[−1.26]	[−1.05]	[−1.77]

Weighted Average Elasticities (Standard Error)—Price: −0.63 (0.04)—Income: 0.46 (0.16)

1978–85: 13 Observations (2SLS)	1st month	2nd month	3rd month	4th month	Total
Weighted Mean (in Percents)	−1.0	−4.8	−4.7	−0.3	−10.7
Mean (in Percents)	−1.3	−4.9	−1.0	5.3	−1.8
(*t*-statistic)	(−0.09)	(−0.18)	(−0.06)	(0.02)	(−0.06)
[*z*-statistic]	[−0.36]	[−0.93]	[−0.88]	[−0.21]	[−0.64]

Weighted Average Elasticities (Standard Error)—Price: −0.50 (0.20)—Income: 1.38 (0.37)

fects and the weighted average where the weights are again the inverse of the variance of each estimated effect. The latter approach seems to yield more reasonable results.[32]

The results are consistent throughout the pre-deregulation period. Accidents, even the most catastrophic ones, appear to have very small average effects on the demand that the airline faces.[33] None of these estimates is statistically different from zero. Even the most negative estimates are quite small relative to the fluctuation caused by other factors. As a point of comparison, the peak-to-trough seasonal variation in demand (with February being the trough and July or August the peak for most carriers) is estimated to average more than 30 percent during these periods. Furthermore, the standard errors of these demand regressions, which represent the random demand component, range from 3 to 31 percent and average 9 percent. It seems clear that prior to deregulation, the effect of a crash on demand was quite small and was very difficult to distinguish from other factors that cause demand fluctuations.

---

[32] The weighted and unweighted average percentage demand deviations are obtained by taking the exponential of the δ parameter estimates. The t- and z-statistics are calculated directly from the δ estimates. They are changed only slightly when they are calculated from the variances of the transformed estimates approximated from a first-order Taylor series.

Since deregulation, consumers' responses to crashes appear to have increased. These results must be qualified, however, because they are based on only 13 accidents, a small sample that may not be representative of the "typical" crash. Though the estimated effects are generally not significant at conventional levels, the results indicate a pattern of negative response to crashes that tapers off after approximately two months.[34] The total loss of demand over the four-month period due to an accident is estimated to (weighted) average 10.7 percent of one month's traffic when estimated by two-stage least squares. When the deregulation period demand regressions are estimated by OLS, assuming that price is exogenous, the estimated (weighted) average quantity effect of a crash is 15.3 percent of one month's demand. The t-statistics for these estimates are not significant, due in part to the large variance in demand for a few carriers. The z-statistics for the 2SLS and OLS estimates are -0.64 and -1.77, respectively, the latter being significantly different from zero at the 10 percent level, using a one-tailed test.[35]

The danger of inferring a systematic demand response from these 13 observations is highlighted by a closer look at the accident of Air Florida in January, 1982. For many reasons, this crash of a Boeing 737 in Washington, D.C., that killed 74 of the 79 people on-board seems to be among the most likely candidates for significant adverse consumer reaction. Air Florida was a new airline without an established safety record. The accident was quickly and conclusively deemed to be the result of pilot error. Furthermore, the error was probably related to inadequate pilot training by the airline regarding flight procedures in subfreezing temperatures. Due in part to its location and the heroic rescues that took place, the crash received extensive media coverage, including mention during the President's state of the union address two weeks later. Finally, most of Air Florida's traffic was carried on the highly competitive routes between northeastern cities and Florida, so travelers could fairly easily switch airlines. Still, the cumulative abnormal demand in January through April of 1982 is estimated to be *positive*, though not significantly so. Despite a slumping economy and virtually the same prices as a year earlier, Air Florida carried 30 percent more passengers in the first quarter of 1982 than in the first quarter of 1981.[36]

---

[33] The results are similar when the analysis includes only accidents in which there were 30 or more fatalities, as well as when the crew-only accidents are excluded.

[34] Though the 1979 American Airlines DC-10 crash is in this group, it does not dominate the analysis. The estimated demand loss for that accident is about equal to the average for the sample of 13.

[35] Due to the question of the effectiveness of the cost index as an instrument in the 2SLS estimation, the demand equations were also estimated restricting the price elasticities to be alternatively, -0.5 and -2.0 and using the actual price as the right-hand side variable. The results are:

Price Elasticity	Weighted Total	Total	t-statistic	z-statistic
-0.5	-12.0 percent	-1.2 percent	-0.25	-0.96
-2.0	-23.2 percent	-16.0 percent	-1.49	-2.32

Total Effect is as a proportion of one-month demand.

[36] Air Florida did have lower-load factors in the first quarter of 1982 than a year earlier, but this appears to be due to an expansion of capacity that occurred at the end of 1981. The load factor for February and March 1982 was above that for December 1981, before the

TABLE 6—ANALYSIS OF THE CAUSES OF DEMAND CHANGE

Descriptive Statistics

Variable	N	Mean	Standard Deviation	Minimum	Maximum
δ_0	91	−0.0155	0.118	−0.500	0.294
δ_1	91	−0.0237	0.137	−0.766	0.348
δ_2	91	−0.0004	0.133	−0.483	0.538
δ_3	91	0.0170	0.152	−0.273	0.924
FATAL	91	46.0659	53.155	1.000	327.000
ART	91	5.3846	6.555	0.000	38.000
PAGE1	91	0.9890	1.111	0.000	7.000
RPMLAG (billion)	91	0.7420	0.794	0.003	3.206
DATE	91	15.7582	9.294	1.000	31.000
FATSUM	91	201.1319	108.998	0.000	454.000
FATSUMOWN	91	11.3297	28.542	0.000	177.000
FAULT	74	0.7703	0.424	0.000	1.000

Observations: 91 Dependent Variable:	δ_0	δ_1	δ_2	δ_3	(×1000)
INTERCEPT	−4.176	20.885	24.174	5.791	
	(28.340)	(33.600)	(31.529)	(30.164)	
FATAL	−0.088	0.039	−0.183	−0.259	
	(0.252)	(0.300)	(0.280)	(0.266)	
ART	1.694	3.154	5.903[a]	2.246	
	(2.372)	(2.833)	(2.646)	(2.503)	
PAGE1	−8.146	16.656	−23.162	6.969	
	(11.967)	(14.823)	(13.877)	(13.063)	
RPMLAG	11.386	−15.072	−13.121	−10.627	
	(13.534)	(16.064)	(15.0661)	(14.589)	
DATE	0.097	−0.332	0.683	0.936	
	(0.852)	(1.004)	(0.940)	(0.905)	
FATSUM	−0.038	−0.084	−0.133	−0.107	
	(0.090)	(0.106)	(0.099)	(0.094)	
FATSUMOWN	0.323	0.043	−0.042	0.130	
	(0.325)	(0.362)	(0.328)	(0.308)	
R-Squared	0.0285	0.0422	0.1100	0.0716	
Adjusted R-Squared	−0.0534	−0.0386	0.0350	−0.0067	
$F(8,83)$	0.348	0.522	1.466	0.915	
P-value	0.9287	0.8167	0.1901	0.5003	

[a] Significantly different from zero at the 10 percent level.

C. *Explaining the Change in Demand*

The regressions of (log) proportional abnormal demand on factors that plausibly affect the magnitude of the crash effect are reported in Table 6.[37] The only statistically significant estimate, the effect of the number of articles on δ_2, is of the "wrong" sign, indicating that greater news coverage damp-

---

crash and historically a month of high demand for north–south travel. January load factor was 1 percent below December. Industry output in the first quarter of 1982 was also suppressed somewhat by the flight restrictions that followed the strike and eventual firing of more than half of the nation's air traffic controllers. It is unclear how much this affected Air Florida's routes.

[37] These regressions are corrected for heteroskedasticity by dividing all variables by the standard error of the estimated δ_k, which is the dependent variable. The results are equally insignificant when this correction is omitted.

ens the negative impact of an accident on demand. This seems to be an aberration, however, and *ART* is not significantly correlated with δ_2. In fact, none of the explanatory variables is significantly correlated with the total (proportional) demand effect of crashes.[38]

The *FAULT* variable is not included in the Table 6 regressions, because we were able to reliably classify only 74 of the 91 accidents as being the airline's fault or not. When the regressions are run with only these 74 accidents and *FAULT* is included, the estimates of its effect are positive—indicating smaller demand reactions when the carrier is at fault—but the associated t-statistics are all less than 1.

As discussed in Section II, significant updating of prior beliefs may be much more likely to result from multiple accidents by the same airline than from a single accident. Yet, neither *FATSUM* nor *FATSUMOWN* are significant in the regressions.[39] These variables indicate a run of accidents for the industry and the firm, respectively.[40]

These insignificant results are not surprising for the pre-deregulation crashes, since there appears to be virtually no demand reaction in the first place. The 13 observations in the post-deregulation period alone are too few to get meaningful results from a regression with eight right-hand side variables. For these 13 observations, however, none of the potential explanatory variables is significantly correlated with any of the δ_k's.

Greater demand losses since deregulation are consistent with the more competitive environment that now exists. Prior to deregulation, carriers faced less actual competition and virtually no potential competition. In the 1978–85 period, consumers had more airlines to choose from than in earlier years and could respond more easily to a crash by switching carriers. Such a substitution effect, if it existed, would manifest itself as a positive change in the demand for competing carriers. Our analysis of the traffic of other carriers following a crash gives weak indication of such an effect, as discussed in Section V, Part E below. Such a substitution effect might become apparent from a more detailed route-by-route study. Nevertheless, our estimates suggest that the revenue implications of such a route-specific response would probably be small relative to the size of a major airline.

D. *Explaining the Change in Shareholder Wealth*

Table 7 presents an attempt to relate the loss in firm value to variables expected to influence the magnitude of that loss. *VALCH1* is the change in the total value of outstanding stock on the first day of the event. *VALCH12* is the cumulative change for the first and second day.[41]

Direct costs—uninsured tort liability, lost equipment value, and expected changes in future insurance rates—are represented by the variable *INJ*, which is the total number of people killed or seriously injured in the accident. The number of passengers seriously injured is included based on the observation that a tort settlement or award for a permanently injured person is often as high or higher than for a wrongful death suit.[42]

---

[38] The insignificant F-statistics for each of the regression in Table 6 further indicate that the insignificant t-statistics are not due to multicollinearity. Though fatalities, articles, and page 1 coverage are highly correlated with one another, none is consistently significant if the regression is run omitting the other two.

[39] Nor is either variable correlated with the proportional change in demands.

[40] Actually, these variables are the cumulative number of fatalities for the industry and the firm in the 365 days preceding the accident. Substitution of the cumulative number of fatal accidents also yields insignificant results.

[41] These variables are constructed by multiplying the percentage change in the firm's stock price on the day of the accident (or the next trading day, if the accident occurred after 3 P.M. New York time or on a day on which the stock exchanges were closed) and the succeeding day by the value of all of a firm's outstanding common stock on the day before the accident.

[42] The capacity of the aircraft, *CAP*, was included as well in early specifications of the regression. *CAP* was intended to reflect the cost, direct and through experience rated insurance, of losing the aircraft. The capacity, however, added very little to the explanatory power of the regression and, because of its high correlation

TABLE 7—ANALYSIS OF THE CAUSES OF FIRM VALUE CHANGE

Descriptive Statistics

Variable	N	Mean	Standard Deviation	Minimum	Maximum
VALCH1 (\$mil)	67	−5.78	27.08	−103.81	61.45
VALCH12 (\$mil)	67	−6.54	36.53	−168.91	86.67
REVCHACT (\$mil)	67	−18.04	132.66	−732.97	458.42
REVCHFIT (\$mil)	67	−12.02	53.10	−243.58	126.81
INJ	67	55.52	63.70	1	382

Observations: 67

Dependent Variable:	*VALCH1*	*VALCH1*	*VALCH12*	*VALCH12*
INTERCEPT	0.200	0.205	0.560	0.585
	(0.419)	(0.417)	(0.651)	(0.647)
INJ	-0.038^{a}	-0.038^{a}	-0.037^{a}	-0.037^{a}
	(0.011)	(0.011)	(0.017)	(0.017)
REVCHACT	−0.001		0.018	
	(0.016)		(0.025)	
REVCHFIT		0.020		0.062
		(0.036)		(0.056)
R-Squared	0.1500	0.1539	0.0757	0.0858
Adjusted R-Squared	0.1234	0.1274	0.0468	0.0572
$F(3,64)$	5.647^{a}	5.819^{a}	2.621^{b}	3.003^{b}
P-value	0.0055	0.0048	0.0806	0.0567

Correlation Matrix

	VALCH1	*VALCH12*	*INJ*	*REVCHACT*	*REVCHFIT*
VALCH1	1.00000	0.82781^{a}	-0.21830^{b}	−0.01945	0.02186
VALCH12		1.00000	-0.22258^{b}	−0.01894	0.04276
INJ			1.00000	0.03043	0.05051
REVCHACT				1.00000	0.43743^{a}

[a] Significantly different from zero at the 5 percent level.
[b] Significantly different from zero at the 10 percent level.

An estimate of indirect costs, that is, lost revenue, was derived from the estimated abnormal demand that the airline experienced in the months following the crash. If, in fact, airlines adapt to lost demand in the short run, by canceling some flights for instance, then the revenue loss will overstate the lost profits due to demand loss. On the other hand, if the effect were to persist beyond the three- to four-month period that is posited here, the revenue loss in this period may understate lost profits. Still, one would expect the coefficient on *REVCH* to be in the neighborhood of 1, a \$1 drop in revenue from a short-run demand loss will generate roughly a \$1 loss in profits.

The change in firm value was regressed on the estimated lost future revenue and the number of fatalities plus serious injuries.[43]

---

with *INJ*, increased the standard error of the coefficient estimate on *INJ* substantially. Therefore, the regressions presented here exclude *CAP*. The estimated effect of *INJ* should be interpreted as the impact of both tort liability and lost equipment, as well as the changes in the cost of future insurance against these losses. The multicollinearity in this relatively small sample does not allow one to separate these effects.

[43] Because the volatility of airline stocks differs quite a bit among airlines, the residuals in such a regression would be expected to be heteroskedastic. The variables

The sample for these regressions includes 67 accidents, the intersection of the observations used in the estimation of demand changes, and the observations available from the *CRSP* data base for equity value changes. Because the regressions of demand impact on causal factors exhibit very poor fits, we tried using alternatively the "actual predicted" revenue impact of each crash (*REVCHACT*), based on δ_{jlk}, as well as the "fitted predicted" revenue impact of each crash (*REVCHFIT*), based on $\hat{\delta}_{jlk}$, in the regressions to explain changes in the equity value of firms due to accidents.

Table 7 shows that neither *REVCHACT* nor *REVCHFIT* is significantly related to the change in firm value. Though the parameter estimates are mostly positive, they are very small—indicating a decrease of between zero and six cents in profits for every decrease of $1 in revenue—and statistically insignificant.[44] The simple correlation between each of these variables and the change in equity value is also small and not significantly different from zero. The number of fatalities plus serious injuries, *INJ*, does appear to be a significant cause of the change in equity value. The parameter estimate in each of the *VALCH1* regressions is significant at the 1 percent level. The parameter estimates for this variable indicates an equity loss of about $38,000 per life.[45]

E. *Externality Effects on Other Airlines*

One airline's accident could affect other airlines through many different transmission mechanisms. The public could react to an airline's crash by deciding that all air travel is less safe, or they could view the information as firm-specific and switch to other carriers. In response to an accident, the FAA could make a change in regulations that would increase production costs for all airlines, or it could redeploy its limited monitoring resources to scrutinize one airline more closely, thereby lessening surveillance of other carriers. The direct costs of a crash could increase the probability that an airline will go out of business, thereby increasing future sales for other carriers.

In both the pre- and post-deregulation periods, Table 8 shows very little evidence that the demand faced by one airline is affected by another carrier's accident. For 1960–77, the average estimated abnormal demand for other carriers after an accident is just slightly negative in the month of a crash and the following month. Neither estimate is significantly different from zero and a 95 percent confidence interval for each estimate of demand change as a fraction of monthly demand is wholly contained within (−2 percent, +2 percent). The results are similarly small in the post-deregulation period, though somewhat less tightly estimated, probably because the sample contains only 13 accidents. Isolation of the eight accidents in which 100 or more people were killed yields weak evidence of a small positive externality during the month of the accident, though it appears to be offset somewhat in the following month.

When one isolates the ten cases in which crashes were estimated to have had the largest proportional own-firm demand effects, however, the average effect on other firms is significant and negative, −1.4 percent in the month of the crash and −0.5

---

for each observation are divided by the square root of the product of the mean squared error from the stock market model regressions (equation (1)) and the value of firm's outstanding equity on the day before the accident. This is not an exact correction. The true standard error of the predicted values includes terms of order 1/N that are ignored in this adjustment.

[44] The descriptive statistics in Table 7 indicate an average *REVCHACT* of −$18 million. This reasonably large negative mean results entirely from the deregulation-era accidents. The mean *REVCHACT* for all of the 78 pre-deregulation accidents for which demand effects were estimated is +$0.3 million, while the mean for the 13 post-deregulation crashes is −$91.1 million. Still, *REVCHACT* is uncorrelated with the value change in the deregulation era. All *REVCH* values for the post-deregulation accidents are calculated using the δ_k values from the 2SLS estimation of the demand functions.

[45] The financial regressions were also run eliminating "crew-only" accidents. The results were not altered qualitatively. The effect of *INJ* on *VALCH1* is estimated to be about $37,000 and significant, while the *REVCH* variables are estimated to have a positive, but insignificant effect.

TABLE 8—ESTIMATED IMPACT OF ACCIDENTS ON OTHER AIRLINES' DEMANDS

1960–77: 78 Accidents, 1116 Observations	1st month	2nd month
Weighted Mean Effect (in Percent)	−0.380	−0.235
Mean External Demand Effect (in Percent)	0.012	0.350
(t-statistic)	(−0.874)	(0.412)
[z-statistic]	[−1.369]	[−0.372]
1978–85: 13 Accidents, 168 Observations	1st month	2nd month
Weighted Mean Effect (in Percent)	−0.809	−1.370
Mean External Demand Effect (in Percent)	1.860	2.748
(t-statistic)	(−0.319)	(0.023)
[z-statistic]	[−0.789]	[−1.005]
8 Accidents with More Than 100 Fatalities, 140 Observations	1st month	2nd month
Weighted Mean Effect (in Percent)	0.827	−0.656
Mean External Demand Effect (in Percent)	1.578	−0.452
(t-statistic)	(0.866)	(−0.359)
[z-statistic]	[1.531]	[−1.032]
10 Accidents with Largest Own-Demand Effects, 176 Observations	1st month	2nd month
Weighted Mean Effect (in Percent)	−1.440	−0.507
Mean External Demand Effect (in Percent)	0.692	2.951
(t-statistic)	(−0.913)	(0.854)
[z-statistic]	[−2.262][a]	[0.747]

Note: Effect of accidents during 1978–85 calculated from two-stage least squares regressions.
[a] Significantly different from zero at the 5 percent level.
[b] Significantly different from zero at the 10 percent level.

percent in the following month.[46] One explanation for this result is that the crashes that have large own-firm effects also impose significant negative externalities on all other firms. An alternative explanation, however, is that both the own-firm and other-firm estimates of demand effects are being influenced by external factors unrelated to the accident that our demand regressions have not fully accounted for.[47] Without including such variables in the demand regressions, which would be quite difficult, it is not possible to distinguish between these hypotheses from the demand data alone. Some information is provided, however, by examination of stock price behavior on crash days of uninvolved airlines.

These ten events produce 160 observations on the movement of other airlines' stock prices. On the first trading day associated with the accident the average equity value change of the other carriers is −0.61 percent with a t-statistic of −1.70, a z-statistic of −2.29, and 63.1 percent negative stock movements. The average second-day movement for other carriers is positive, +0.44 percent, with a t-statistic of 1.24, a 1.48 z-statistic, and 43.8 percent negative.[48]

---

[46] These 10 cases represent 11 crashes, because in one of these cases 2 crashes occurred in the same month.

[47] These could include supply restrictions such as might be due to fuel rationing, a shortage of air traffic controllers, or limited short-run availability of aircraft. External demand factors, such as the prices of alternative modes of transportation or the presence or absence of demand stimulating incentives, for example, frequent flyer programs, could also have such industrywide effects.

[48] By contrast, the own-firm effects for these 10 events were

Day of Event: AR −1.29 percent t-stat −1.44
z-stat −1.44 70 percent neg.
Day after event: AR 0.11 percent t-stat 0.12
z-stat 0.53 40 percent neg.

Though the first-day movements suggest some spillover effects, the second day appears to offset this effect. Further, if there are spillover effects, they are not associated with the largest crashes. For the eight events with 100 or more fatalities, the movement of other airlines' stocks averages virtually zero: +0.05 percent on the day of the event, −0.06 percent on the day following the event, neither figure at all significant.

VI. Conclusion

This paper has examined the losses suffered by airlines in connection with a crash. On average, crashes are associated with a statistically significant loss in equity value of 1 percent or $4.5 million. Since most crashes in the sample involve total destruction of the aircraft and an average of more than 40 fatalities, the average firm loss appears to be below the total social costs of the accident. This is consistent with the practice of carrying insurance against the direct costs of major crashes, and indicates that the insurance cost is only partially dependent on an airline's own experience.

The investigation of consumer response found that before deregulation, travelers did not respond to crashes to an extent that is statistically discernible. In the period since deregulation, the consumer response appears to have increased. The reaction to the 13 post-deregulation crashes that we studied is weakly significant. Still, the sample of crashes in this period is quite small and may be idiosyncratic.[49]

Previous work has speculated that the wealth losses associated with accidents or product recalls result from adverse consumer reaction. In the case of airline accidents, our direct estimation of consumer response does not support this hypothesis. The on-site airline safety supervision by the government may explain why the effects found here are quite a bit smaller than the effects due to prescription drug recalls found by Jarrell and Peltzman (1985). Alternatively, the greater consumer familiarity with airline travel, as compared to new drugs, may explain why consumers seem to infer much more information from drug recalls than from an airline crash.

Chalk's work suggests that air carriers may infer more information about aircraft manufacturers from aircraft-related crashes than consumers derive about airlines from their accidents. Two factors may explain this difference. First, airlines are experts in the purchase of aircraft and thus may be better equipped than consumers to process the information from an accident. Second, there may be more information in aircraft-related accidents about other aircraft from the same manufacturer than there is in airline accidents in general (or even those deemed to be the carrier's fault) about other flights by the same airline. That is, accidents not caused by aircraft failure may be more idiosyncratic.

It must be stressed that our results and conclusions are conditional upon the level of FAA monitoring and the outstanding safety record that the industry maintains. The relatively small demand effects observed, whether due to very limited updating of prior beliefs or to a low *marginal* valuation of safety, would probably increase if accident records deteriorated substantially.

DATA APPENDIX

The following data series were used in the statistical calculations (All monetary variables are expressed in January 1985 constant dollars.):

Revenue Passenger-Miles—monthly, by carrier, (a) domestic scheduled service, (b) international scheduled service, *Source:* Air Carrier Traffic Statistics, 1957–1985, CAB.

Revenue Passenger-Miles—monthly, industry total, scheduled plus nonscheduled services. *Source:* Air Carrier Traffic Statistics, 1957–1985, CAB.

Available Ton-Miles—monthly, total certificated carriers, (a) domestic scheduled services, (b) international scheduled services. *Source:* Air Carrier Traffic Statistics, 1957–1985, CAB.

Total Revenues—quarterly, by carrier, (a) domestic scheduled service, (b) international scheduled service.

---

[49] Recent press reports support a skeptical view of the post-deregulation results. Delta Airlines suffered few cancellations following a spate of near accidents in 1987, almost all of which were due to errors by flight crews or maintenance personnel. See *New York Times*, July 25, 1987. Vocal dissatisfaction with an airline due to poor performance in other areas also does not seem to indicate much avoidance of the carrier. See *Wall Street Journal*, November 19, 1987.

Source: Air Carrier Financial Statistics, 1957–1985, CAB.

Yield—monthly, by carrier, (a) domestic scheduled services, (b) international scheduled services. **Construction:** Total Revenues divided by Revenue Passenger-Miles, which have been aggregated to be quarterly data. This produces quarterly yield. Monthly yield is the produced through linear interpolation. Each yield number is assumed to correctly reflect the second month in each quarter. The yield for the first month of quarter i is then $= 0.667 * yield_i + 0.333 * yield_{i-1}$, where the subscripts indicate quarters. The yield for the third month of quarter i is then $= 0.667 * yield_i + 0.333 * yield_{i+1}$.

Industry Cost Index—monthly, total certificated carriers, (a) domestic scheduled services, (b) international scheduled services. *Source:* Air Carrier Financial Statistics, 1957–1985, CAB. Air Carrier Traffic Statistics, 1957–1985, CAB. **Construction:** Total Expenses = line 29–line 26–line 23–line 21 all from quarterly air carrier financial stats. Available Ton Miles = line 5 of air carrier traffic stats aggregated for the 3 months in each quarter. This produces a quarterly index, which is then smoothed to monthly using linear interpolation, as with yields above.

Firm Fatalities—by carrier. *Source: Briefs of Accidents: U.S. Civil Aviation*, 1958–85.

Total Industry Fatalities. *Source: Briefs of Accidents: U.S. Civil Aviation*, 1958–85.

Cause of Accident. *Source: Briefs of Accidents: U.S. Civil Aviation*, 1958–85.

News Coverage, Total Articles—by accident. Total number of articles. *Source: New York Times* and *Chicago Tribune* Index of Articles, 1959–1985.

News Coverage, Total Days of Page-One Exposure—by accident. Total number of articles. *Source: New York Times* and *Chicago Tribune* Index of Articles, 1959–1985.

REFERENCES

Borenstein, Severin, *Price Discrimination in Free-Entry Markets*, unpublished doctoral dissertation, Cambridge, MA: MIT, 1983.

Chalk, Andrew, "Market Forces and Aircraft Safety: The Case of the DC-10," *Economic Inquiry*, January 1986, *25*, 43–60.

______, "Market Forces and Commercial Aircraft Safety," *Journal of Industrial Economics*, September 1987, *36*, 61–81.

Chance, Don M. and Ferris, Stephen P., "The Effect of Aviation Disasters on the Air Transport Industry," *Journal of Transport Economics and Policy*, May 1987, *21*, 151–65.

Crafton, Steven M., Hoffer, George E. and Reilly, Robert J., "Testing the Impact of Recalls on the Demand for Automobiles," *Economic Inquiry*, October 1981, *19*, 694–703.

Fama, Eugene F., Fisher, Lawrence, Jensen, Michael C. and Roll, Richard, "The Adjustments of Stock Prices to New Information," *International Economic Review*, February 1969, *10*, 1–21.

Graham, David R., Kaplan, Daniel P and Sibley, David S., "Efficiency and Competition in the Airline Industry," *Bell Journal of Economics*, Spring 1983, *14*, 118–38.

Hoffer, George E., Pruitt, Stephen W. and Reilly, Robert, J., "The Impact of Product Recalls on the Wealth of Sellers: A Reexamination," *Journal of Political Economy*, June 1988, *96*, 663–70.

Ippolito, Richard A., "Estimating Airline Demand with Quality of Service Variables," *Journal of Transport Economics and Policy*, January 1981, *17*, 7–16.

Jarrell, Gregg and Peltzman, Sam, "The Impact of Product Recalls on the Wealth of Sellers," *Journal of Political Economy*, June 1985, *93*, 512–36.

Mitchell, Mark L. and Maloney, Michael T., "The Role of Market Forces in Promoting Air Travel Safety," unpublished manuscript, Center for Policy Studies, Clemson University, May 1988.

Morrison, Steven A. and Winston, Clifford, *The Economic Effects of Airline Deregulation*, Washington: The Brookings Institution, 1986.

Reilly, Robert J. and Hoffer, George E., "Will Retarding the Information Flow on Automobile Recalls Affect Consumer Demand?," *Economic Inquiry*, July 1983, *21*, 444–47.

Rose, Nancy L., "Profitability and Product Quality: Economic Determinants of Airline Safety Performance," Sloan School of Management Working Paper No. 2032-88, MIT, June 1988.

Salinger, Michael A., "Event Studies When the Event Changes the Market Model Parameters," First Boston Working Paper No. FB-88-02, Graduate School of Business, Columbia University, New York, January 1988.

Chicago Tribune, *Chicago Tribune Index*, Wooster, OH: Bell & Howell, Co., 1959–1981.

New York Times, *The New York Times Index*, New York: New York Times, 1959–1985.

"Delta Offers Refunds on Safety Issue," *New York Times*, July 25, 1987, *137*, 37–38.

U.S. Civil Aeronautics Board, *Air Carrier Traffic Statistics*, Washington: USCAB, 1957–1985.

U.S. Civil Aeronautics Board, *Air Carrier Financial Statistics*, Washington: USCAB, 1957–1985.

U.S. Department of Transportation, National Transportation Safety Board, *Briefs of Accidents*: *U.S. Civil Aviation*, Washington: USDOT, 1958–1985.

"Gluttons for Punishment? Fliers Continue Using Airlines They Hate," *Wall Street Journal*, November 19, 1987, *69*, 33.

Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing

By B. PETER PASHIGIAN*

The frequency of sales has increased dramatically in recent years. Paradoxically, the percentage markup has also increased. Markdowns and markups are higher in some departments of department stores than in others. The major theme of the paper is that the growing role of fashion and product variety is an important reason for these increases and for the differences between merchandise groups.

Markdowns and sales have long been ubiquitous in retailing. They occur frequently but they have not been studied systematically by economists. Recently renewed interest in the subject has been expressed by two groups: economic theorists Hal R. Varian (1980), Steven Salop and Joe E. Stiglitz (1982), and Edward P. Lazear (1986), and marketing specialists, who have concentrated on the growing use of promotional price deals in package goods.

This paper expands on the topic in two ways. It extends Lazear's theory of clearance sales to allow for industry equilibrium. This is a useful extension because it shows how the equilibrium percentage markup is jointly determined along with the percentage markdown and the percentage of goods sold on sale and why these three measures might be expected to either increase or decrease together. More importantly, the paper takes the next step. It confronts the theory of clearance sales with some evidence and determines just how well the theory can explain some interesting time-series and cross-sectional regularities in sales offered by department stores. The empirical section includes an in-depth and reasonably comprehensive study of actual sales. The percentage markup and dollar markdowns taken by department stores relative to dollar revenue are traced over an extended time, from 1925 to 1984, and then compared across merchandising groups at several points in time. The most striking feature of the time-series is the increase in the percentage markup and the frequency of sales since 1970. These increases can be properly viewed as revolutionary since neither series exhibited any clear trend during the previous 40 to 45 years. A considerable part of the modeling effort is designed to discover why these increases began around 1970 and why there are significant but changing differences in markups and markdowns between merchandise groups. A major theme of the paper is that these recent increases and differences between groups are due to the growing importance of fashion in merchandising. The term *fashion* is used synonymously with variety throughout this paper and refers to greater (price) uncertainty facing stores about the future popularity of colors, patterns, and fabrics. To avoid any misunderstanding, the term "fashion" is not meant to refer solely to "high" fashion clothing.

---

*Graduate School of Business, University of Chicago, 1101 E. 58th Street, Chicago, Illinois 60637. This research was supported by the Graduate School of Business, University of Chicago and the Center for the Study of the Economy and the State. The author acknowledges helpful comments from Allan Hunter, Edward Lazear, Carl Priestland, Scott Neslin, Stephen Spurr, George Stigler, Arnold Zellner, and participants at seminars presented at the University of Chicago, Clemson University, Duke University, Federal Trade Commission, Massachusetts Institute of Technology, Northwestern University, and Rochester University. Numerous helpful conversations were held with knowledgeable industry officials and are gratefully acknowledged. While most industry officials acknowledge the role of fashion, some may place greater importance on the growth of store competition. Responsibility for error rests with the author. Peter Allan Pashigian assisted with the graphics. Will Carrington, Brooks Pierce, Ken Troske, and Chinhui Juhn served as able research assistants.

Section I extends Lazear's theory of clearance sales to allow for industry equilibrium. The effect of increased dispersion in reservation prices on the optimal percentage markup, the percentage markdown, and the percentage of goods sold on markdown is shown for several price distributions. Implications of the fashion hypothesis are derived. Section II examines the time-series behavior of markups and markdowns. Time-series changes in the use of colors and types of clothing are documented to show increases in the demand for fashion. The estimated effects of different proxy variables for fashion on percentage markup and on markdowns are presented. Section III shows how percentage markups and markdowns have changed over time for women's fashion, women's standard, men's, teens', and infant apparel groups. Some of these changes appear to be caused by the growing importance and spreading influence of fashion in apparel merchandising. The paper ends with a summary.

I. The Theory of Clearance Sales

Sales can be conveniently classified as preseason, within-season, and clearance sales. Of the three, clearance sales are the more familiar and somewhat easier to understand. They occur because style or color or pattern changes are so difficult to predict. Lazear's theory of clearance sales features one form of uncertainty—what customers will pay for individual dresses within a line of dresses. His model considers a risk-neutral store that has already purchased the line but is uncertain about which color will be popular in the coming season and could be sold for higher prices. The prior cumulative price distribution of reservation prices that consumers are willing to pay for the different colored dresses in the line is denoted by $F(P)$. Because the retailer cannot identify which dresses could be sold for high prices and which for low prices, all dresses are offered in the first period at an initial price, $P0$. The dresses with colors for whom consumer reservation prices exceed $P0$ will sell out quickly. If a dress with a particular color did not sell in the initial period, it means that consumers' reservation prices for that color are less than $P0$. The store revises the prior distribution of prices in light of the market evidence and selects an optimal markdown price, $P1$.[1]

The markdown price is selected to maximize expected revenue:

$$(1)\quad R1 = P1\int_{P1}^{P0} f(P)/F(P0)\,dP$$
$$= P1\{F(P0)-F(P1)\}/F(P0)$$

where $P_0 \geq P_1 \cdot f(P)/F(P0)$ is the revised price distribution given that the dress did not sell at $P0$. When selecting the optimal markdown price, the store has a tradeoff to make between a lower markdown price and an increase in the proportion of the remaining dresses that will be sold by the end of the markdown period. Given $P0$, the optimal markdown price satisfies

$$(2)\quad F(P0)-F(P1)-f(P1)P1=0.$$

The optimal first-period price maximizes expected revenue given optimal pricing behavior in the markdown period. Expected revenue in the first period is

$$(3)\quad R0 = P0\{1-F(P0)\}$$
$$+P1\{1-F(P1)/$$
$$F(P0)\}F(P0).$$

For a dress selected at random, expected revenue equals $P0$ times the probability of sale in the initial period plus $P1$ times the probability of selling the dress in the markdown period. It is assumed that a dress that has not sold at the end of the markdown period is given to charity. The optimal initial price will satisfy

---

[1] Lazear's model ignores strategic behavior by consumers. Consumers are less likely to behave strategically when the selling seasons are short, fashions change from season-to-season, and lead times between order and sales are long. Consumers do not wait to purchase a fashion item in the markdown period because the item may be sold out before then and may not long be popular in the next season.

$$(4)\quad (P1-P0)f(P0)+1-F(P0)=0.$$

Lazear used equations (2) and (4) to determine the optimal initial and markdown prices.[2] He focused on the optimal price policy of a store that has *already* ordered a line of clothing. The analysis can be enriched by allowing for industry equilibrium through the actual or threatened entry of firms. New stores will enter or old stores will exit if expected profits differ from zero. So, market equilibrium requires

$$(5)\quad P0\{1-F(P0)\} + P1\{F(P0)-F(P1)\}=C,$$

where C is the per unit (= marginal) cost of purchasing the merchandise.

A market equilibrium exists when equations (2), (4), and (5) are satisfied. The adjustment to this equilibrium is not formally modeled in this paper. A brief narrative might help clarify just how this equilibrium could be reached. Begin by assuming the parameters of the price distribution are exogeneously determined by the tastes of and the information held by consumers. For a given price distribution, each store will post the optimal initial and then the optimal markdown price while satisfying equations (2) and (4). If expected profits are positive at these prices, then new stores will have an incentive to enter and to acquire customers of established stores. Because all stores are assumed to be alike, any new store that enters will face the same price distribution as existing stores. So a new store will select the same prices that satisfy equations (2) and (4). Assume that new stores that post identical prices as established stores cannot attract customers from the established stores except by offering higher quality merchandise, that is, by increasing C. The actual or implied threat of quality upgrading by new stores will force established stores to upgrade quality or lose customers. Competition through quality upgrading will continue until expected profits are competed away.[3]

Under this scenario, equations (2), (4), and (5) can be solved for the endogenous variables, $P0$, $P1$, and C, as functions of the parameters of the price distribution. An interesting question is how $P0$, $P1$, and C change as the range of reservation prices increases. Unambiguous comparative static predictions are not easy to derive without placing more structure on the price distribution. Considerable insight can be gained by considering price distributions that can be written in the form of $F(P/X1,\ X0/X1)$, where $X0$ and $X1$ represent the highest and the lowest (reservation) price that customers are willing to pay for the dresses in the line. The larger the ratio of $X0$ to $X1$ the greater is price uncertainty. Not only can $X0/X1$ change over time but it will be larger in those merchandise groups where there are frequent changes in color, fabric, pattern, and silhouette from season-to-season. Given the assumed form of F, it is possible to solve equations (2) and (4) for $P0/X1$ and $P1/X1$ as functions of $X0/X1$. After substituting these expressions for $P0/X1$ and $P1/X1$ into equation (5), there will be only one value of $C/X1$ that satisfies equation (5) for each value of $X0/X1$. As a result, the equilibrium functional relation between $C/X1$ and $X0/X1$ can be expressed as

$$(6)\qquad C/X1=G(X0/X1).$$

Given $X0/X1$ and the corresponding equilibrium values for $P0$, $P1$, and C (relative to $X1$), it is then possible to calculate

---

[2] The model can be expanded to allow for noise to make the inference problem more difficult.

[3] Obviously, this description is incomplete and needs to be elaborated on if the whole distribution of reservation prices shifts to the right when quality upgrading occurs. Offering higher quality merchandise would change the parameters of the price distribution and shift the price distribution to the right. A zero-expected profit equilibrium can be achieved as long as the distribution of prices does not shift too rapidly to the right as C increases.

TABLE 1–OPTIMAL PRICES, MARKUPS, AND MARKDOWNS FOR UNIFORM AND SYMMETRIC TRIANGULAR DISTRIBUTIONS

$\frac{X0}{X1}$ (1)	$\frac{P0}{X1}$ (2)	$\frac{P1}{X}$ (3)	$\frac{C}{X1}$ (4)	PMD (5)	MPS (6)	PMU (7)	$\frac{F\left(\frac{P0}{X1}\right)-F\left(\frac{P1}{X1}\right)}{1-F\left(\frac{P1}{X1}\right)}$ (8)
			Uniform Distribution				
1.20	1.10	1.00	1.05	.08	.05	.05	.5
1.50	1.25	1.00	1.12	.20	.11	.10	.5
2.00	1.50	1.00	1.25	.33	.20	.17	.5
3.00	2.00	1.00	1.50	.50	.20	.25	.5
4.74	3.16	1.52	2.00	.50	.20	.37	.5
10.92	7.28	3.64	4.00	.50	.20	.45	.5
25.63	17.05	8.57	8.93	.50	.20	.48	.5
			Symmetric Triangular Distribution				
1.40	1.17	1.01	1.11	.13	.05	.05	.35
1.74	1.32	1.05	1.21	.21	.08	.08	.36
2.28	1.57	1.14	1.37	.28	.12	.12	.38
4.07	2.45	1.58	1.95	.39	.17	.21	.41
6.11	3.50	2.16	2.63	.39	.19	.25	.42
8.99	5.00	3.01	3.62	.40	.21	.28	.43
19.01	10.43	6.10	7.09	.41	.23	.32	.46

the equilibrium percentage markdown,

$$\text{PMD} = (P0 - P1)/P0; \tag{7}$$

the equilibrium percentage markup (PMU)

$$\text{PMU} = (P0 - C)/P0, \tag{8}$$

and total dollar markdowns as a percentage of dollar revenue, MPS,

$$\begin{aligned}\text{MPS} = \{P0 - P1\} \\ \times \{F(P0) - F(P1)\}/ \\ [P0\{1 - F(P0)\} \\ + P1\{F(P0) - F(P1)\}].\end{aligned} \tag{9}$$

Table 1 illustrates the equilibrium solution for seven values of $X0/X1$ for two symmetrical price distributions. The upper panel shows the results for a uniform price distribution and the lower panel for a symmetrical triangular distribution. Columns 2, 3, and 4 show the equilibrium values for $P0$, $P1$, and C (relative to $X1$) and columns 5, 6, and 7 show PMD, MPS, and PMU, respectively. Column 8 shows the ratio of the total units sold at the markdown price to total units sold in both periods.

The results are very informative. They indicate the optimal percentage markdown, the percentage markup, and dollar markdowns relative to dollar sales either rise (for the symmetrical triangular distribution) or rise and then stabilize (for the uniform distribution) as $X0/X1$ increases. The most important implication is that the clearance theory predict that fashion goods will have both a higher-percentage markup and a larger ratio of dollar markdowns relative to dollar sales. An increase in price uncertainty will cause an increase in both the percentage markup and dollar markdowns relative to dollar sales. The extended theory of clearance sales provides a parsimonious explanation of why department stores have increased their percentage markup while they have been simultaneously selling more items at markdown prices. The clearance theory can also be helpful in explaining why pricing policies differ across departments in a store. It can explain why the percentage markup and dollar markdowns (relative to sales) will be higher for fashion garments than for nonfashion garments, for example, teenager versus men's tailored clothing. Another interesting result is that $P1$ is less than C. Customers who are indifferent to fashion trends and who purchase in the markdown period are purchasing remnants at a price less than C. These customers purchase at lower prices because of the preference for fashion by the majority of customers. If there were no price uncertainty, all buyers would purchase at marginal cost. A final result is that the fraction of goods sold at markdown will vary with $X0/X1$ but in ways that depend on the form of the price distribution. It is independent of $X0/X1$ for the uniform distribution and increases with $X0/X1$ for the symmetrical triangular distribution.[4]

II. Time-Series Changes in Markdowns and Markups

Figure 1 shows dollar markdowns as a percentage of dollar sales (MPS) for all department stores in the sample and the cumulative markon (PMU) from 1925 to 1984. The cumulative markon is a term used in the industry and is an approximation to PMU.

Dollar markdowns relative to dollar sales and the percentage markon have changed noticeably over three time periods. First, dollar markdowns relative to total revenue rose with the onset of the Great Depression. Markdowns probably increased because the severity of the Depression was grossly underestimated. Second, markdowns were unusually low during World War II. The decline in markdowns during World War II was due to price controls, output restrictions, and rationing. Prices were artificially fixed below equilibrium levels. Merchandise was scarce and so eagerly purchased by customers that markdowns were limited to the few remnants left on shelves.

Third, and most important, has been the large and steady increase in markdowns over the last 15 to 20 years which follows a period of about 40 years of comparative stability. Figure 1 clearly shows department stores have been taking higher-percentage markups in recent years while they have been simultaneously selling relatively more merchandise at markdown prices. The conjoint increase in the percentage markon and in MPS would no doubt appear paradoxical to the untrained eye. If the increase in the percentage markup is due to persistent and large increases in demand, then markdowns should have been applied less frequently. Similarly, if increased competition from discount and off-price stores is the reason why department stores are offering more frequent sales, then why has this competition not limited or reduced the size of the percentage markup? Both the demand or the competition hypothesis would imply an inverse relationship between the change in the percentage markon and the change in MPS. Neither can completely explain why both PMU *and* MPS have been rising.

The theory of clearance sales provides a plausible and parsimonious explanation of why the percentage markon and MPS have both increased. The theory implies that MPS and PMU will rise when uncertainty increases and uncertainty will increase as fashion and product variety become more important in the sales of apparel and other merchandise.[5]

---

[4] Very similar qualitative results were obtained when the price distributions (right triangular) were asymmetric. The one notable difference is the fraction of goods sold at markdown decreased (increased) with increases in $X0/X1$ when the distribution was skewed to the right (left).

[5] Reliance on the price discrimination hypothesis to explain the higher percentage markup and higher markdowns relative to dollar sales should be resisted. It is doubtful whether the price discrimination hypothesis should be applied to an industry as competitive as the retail clothing industry is. Not only do many different types and sizes of outlets sell clothing, for example, department, discount, apparel, specialty, and off-price stores, but firms enter and exit the industry frequently. If the price discrimination hypotheses were to explain the data, they would imply the monopoly power of these different stores has been increasing over time or the difference between the price elasticities of demand of the two groups of customers had increased. The price discrimination hypothesis would also have to explain why MPS has increased more in selected merchandise groups, for example, teenager and junior or men's clothing (see Section III).

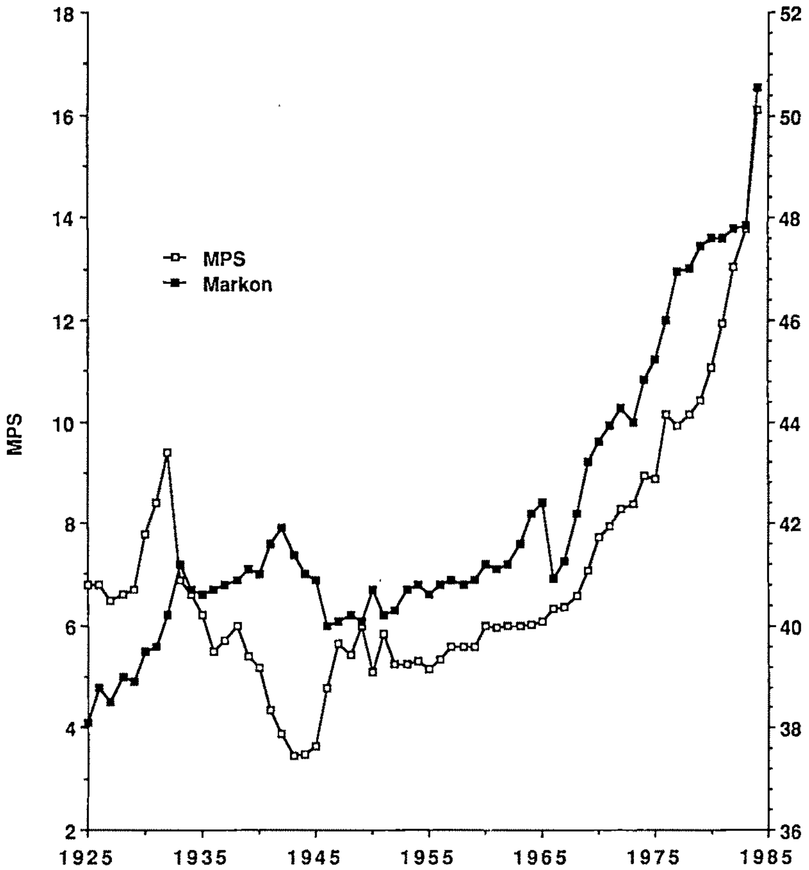


FIGURE 1. MARKDOWNS RELATIVE TO DOLLAR SALES AND PERCENTAGE MARKON, 1925–84 (*Source*: See Data Appendix.)

A. *Illustrations of Fashion Changes*

Some episodic evidence also indicates price uncertainty increased between the mid-1960s and mid-1970s. Color and fashion are intimately linked. Greater use of colors or prints would mirror a shift toward more fashion merchandise. Surprisingly, color use data are scarce and available for only a few products over long periods of time. Data showing the use of whites, prints, and solids are available for bed sheets and to a more limited extent for men's dress shirts. Figure 2 shows the percentage of bed sheet sales (produced by domestic manufacturers) that are white or fancies, which include prints, jacquards, and other special designs (the market share of other solid colors is not shown). The market share of white sheets drops precipitously from about 65 percent in the mid-1960s to about 16 percent by 1975 while the market share of fancies increases from 15 percent in the mid-1960s to 75 percent by the mid-1970s. Over a 10-year span, there was a dramatic increase in the use of prints. Predicting which print patterns and colors will be popular is considerably more difficult than predicting the demand for whites.

A remarkably similar but even more precipitous decline in the demand for whites occurred in the men's dress shirt market. Figure 3 shows the market share of white dress shirts in 1962 was around 72 percent and declined to about 52 percent in 1967

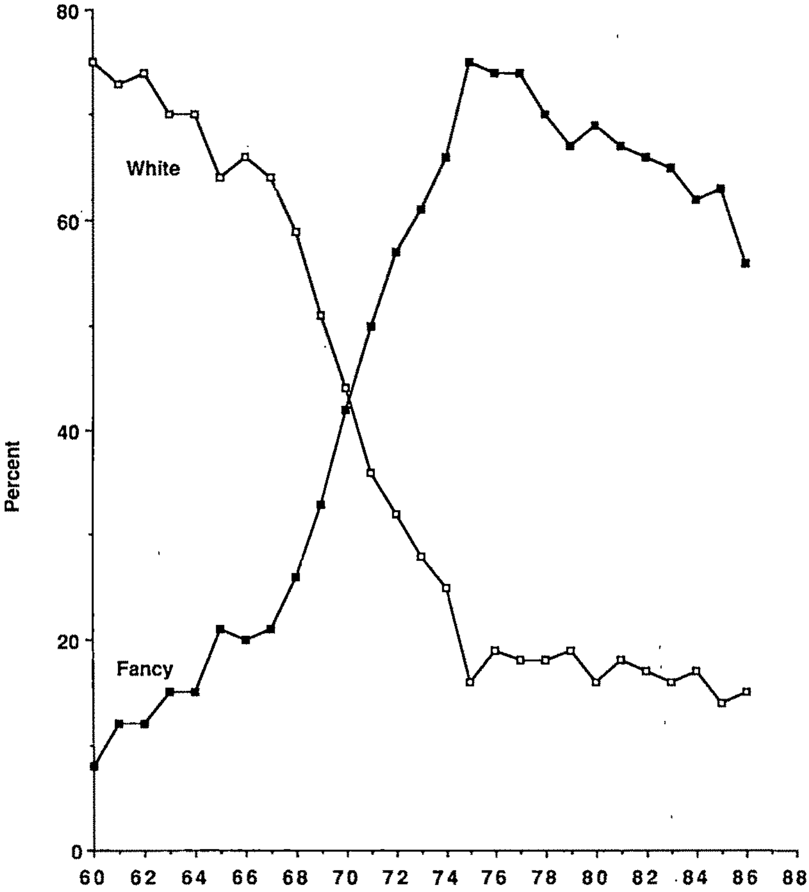


FIGURE 2. MARKET SHARES OF WHITE AND FANCY BED SHEETS, 1960–86
(*Source*: See Data Appendix.)

and to 19 percent in 1970. Complete data for the 1970s are unavailable. However, the market share of white dress shirts probably remained low throughout the 1970s. Data for the eighties are available and show the market share of white shirts is between 19 and 22 percent while the percentage of fancy (patterns, stripes, etc.) is increasing. In this market as well, there has been a shift away from whites to other colors and fancies.

The declining use of more formal and tailored clothing, for which there are accepted guidelines, and the greater reliance on sportswear, for which there are fewer accepted guidelines, has also increased price uncertainty. For women's sportswear, the consumer has greater latitude to mix and coordinate tops with bottoms and to match colors, patterns, and fabrics of blouses, shirts, and sweaters with those of skirts and pants. There are fewer guidelines and standards now and this means that store buyers are even less able to predict which combinations will be matched. The magnitude of the shift toward women's sportswear can be seen in Figure 3. The market share of women's sportswear as a percentage of total dollar sales of women's dresses *and* sportswear is graphed from 1963 to 1982.[6] The market

---

[6] The market shares were calculated from the sales of sportswear and dresses by department stores, women's ready-to-wear stores, and family clothing stores.

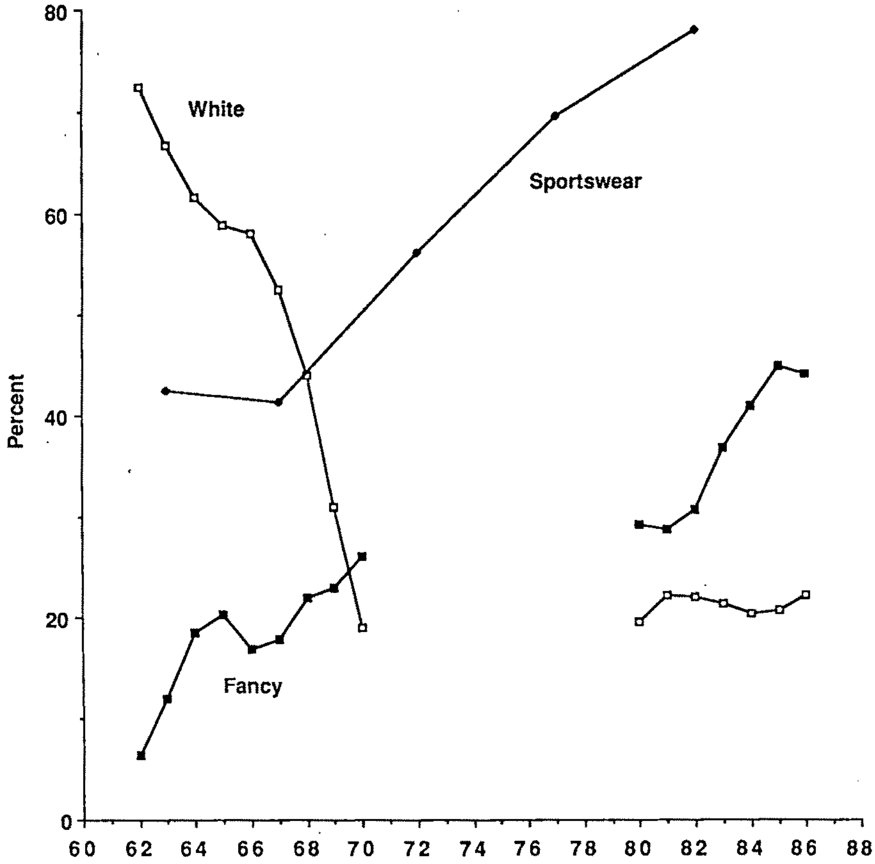


FIGURE 3. MARKET SHARES OF WHITE AND FANCY MEN'S DRESS SHIRTS, 1962–86, AND WOMEN'S SPORTSWEAR, 1963–82 (*Source*: See Data Appendix.)

share of sportswear remained roughly constant between 1963 and 1967, hovering around 41 or 42 percent. Since then, it increased to 56 percent in 1972, 69 percent in 1977, and reached 78 percent in 1982, the last available Census year. In women's apparel (and men's apparel) there has been a dramatic transition toward more casual clothing where there is greater opportunity for individual expression and creativity through product selections.

Although these illustrations are episodic and selective, they do indicate the transition toward greater demand for product diversity began roughly in the mid- to-late 1960s and has continued throughout the 1970s and 1980s.

B. *Regression Results*

Figure 1 indicates the markdown and markup policies of department stores in the last 15 to 20 years were affected by different determinants than in the previous decades. The subsequent statistical analysis of markdowns and markups will concentrate on the period from 1956 to 1984.[7] Markdowns relative to sales and the percentage markup in the period between 1925 and 1955, and especially from 1925 to the end of World War II were more affected by unexpected demand

[7] I selected 1956 as the initial year because clothing import data first became available in that year.

shocks at the onset of the Depression or caused by price and output controls during World War II.[8] The variance in the growth rate of department store sales during the inter-World War II period was larger than in the post-World War II period.

The period from 1956 to 1984 is of greater interest because of the large increases in MPS and PMU. During this period, MPS and PMU appear to reflect the increasing risk associated with predicting styling and fashion changes than the risk associated with predicting aggregate clothing demand. The regression analysis is designed to identify which of a small set of proxies for fashion had a significant effect on MPS and PMU.

The growth in clothing imports relative to domestic shipments of apparel appears to be a major reason for the increase in MPS and PMU. In 1956, clothing imports accounted for slightly more than 1 percent of total shipments of apparel by U.S. apparel plants. By 1984, clothing imports had increased to slightly more than 23 percent of shipments of apparel merchandise. The growth of imported clothing can be expected to increase MPS and PMU for two reasons: Longer lead times are required for ordering foreign merchandise and result in larger forecasting errors when styles and fashions are changing. One consulting firm estimates the mean lead time for a domestic order is 14.2 weeks, for a European order is 21.7 weeks, for an order from a major exporter in the Far East, for example, Hong Kong, is 28.1 weeks and for an order from other Far Eastern countries is 34.7 weeks (U.S. International Trade Commission). Second, the introduction of country-by-country import quotas in the early 1970s induced foreign suppliers to substitute from lower-priced commodity apparel and fabric imports to higher-priced short-run fashion clothing. This quality upgrading has been noted by other authors and was reflected in more rapid increases in imported apparel prices than in domestic apparel prices after the quotas were introduced.[9]

Technological changes may also be indirectly responsible for the increase in MPS. The shuttleless loom is fast replacing the shuttle loom and accounts for a growing percentage of all looms in the United States, in other advanced countries, for example, Italy and even in some countries in the Far East, for example, Taiwan and Hong Kong. Shuttleless looms weave fabric at much faster speeds than do shuttle looms. In the United States the growing use of shuttleless looms has been accompanied by an increasing percentage of shuttleless looms that have a multi-filling insertion capability, which means that multicolor and pattern fabric can be woven. This trend toward looms with a multi-filling insertion implies that more fashionable clothing can be woven. Technical changes in knitting have also reduced the cost of producing short-run fashion fabric.[10] Unfortunately, the data on the percentage of shuttleless looms with the multi-filling insertion capability are not available throughout the 1956–84 period. Consequently, the percentage of shuttleless looms in the United States is used as an imperfect substitute for

---

[8] A statistical analysis of the earlier period indicates the MPS is inversely related to unexpected demand changes. The lower market demand was relative to expected demand, the larger was MPS. These results indicate gross margins fall during large downturns and rise during large upturns in the economy. While gross margins may be pro-cycle, it may take large changes in economic performance to detect this effect.

[9] William R. Cline (1987, p. 173) and Jeffrey S. Arpan et al. (1978, p. 64) show apparel unit values began to increase in 1972 after country import quotas were imposed on Hong Kong, Japan, Korea, and Taiwan. Clothing imports into the United States increased from 27 percent of the sum of clothing plus textile imports in 1956 to 75 percent in 1984. This reflects the shift from lower-priced fabric to higher-priced clothing and apparel imports.

[10] Richard Paul Olsen (1970, pp. 199–200) and Kathryn M. Greenwood and Mary F. Murphy (1978, pp. 134–35) have noted the introduction of electronic knitting machines which permit the knitting of fabric with different patterns and colors. Prior to these technological innovations machines would be down for as much as eight hours when a pattern was changed. Downtimes were reduced drastically with the introduction of electronic knitting machines. Hence, the set-up cost for a pattern change declined with these technological innovations. On the other hand, U.S. textile firms have been criticized for being less willing to produce short-run fashion fabric than have foreign producers.

TABLE 2—DETERMINANTS OF DOLLAR MARKDOWNS RELATIVE TO DOLLAR SALES (MPS) AND PERCENTAGE MARKON, 1956–84 (t-statistics listed below coefficient estimates)

	Dollar Markdowns Relative to Dollar Sales			Markon
	(1)	(2)	(3)	(4)
1. Constant	1.31	1.88	1.13	33.38
	.7	1.7	1.3	14.2
2. Percent Clothing Imports (t)	24.54	16.99	28.07	49.32
	3.5	2.4	4.6	4.9
3. Percent Clothing Imports ($t-1$)			−14.24	
			−1.9	
4. Percent Shuttleless Looms (t)	.13	.07		−.07
	3.2	1.5		−1.3
5. Percent of Population 10- to 34-years old (t)	11.16	2.74		18.94
	2.1	.4		2.8
6. Markdown ($t-1$)		.46	.77	
		2.5	4.4	
R^2	.979	.989	.988	.978
$\sqrt{MSE}$	.303	.297	.304	.430
Rho (ρ)	.39			.06

Note: For definition of variables, see Data Appendix.

the growing capacity to produce fashion merchandise.

Finally, the fraction of the population between 10 and 34 may be directly related to MPS. Younger members of the population appear to be more fashion conscious. This finding should, however, be considered with caution. The fraction of the population between 10 and 34 was not found to be a significant determinant of MPS over the 1925–55 period even though this fraction of the population declined appreciably from 1940 to 1955. During the 1956–84 period, this fraction has been rising except for the last few years.

Regression results are presented in Table 2. In columns 1–3 the dependent variable is MPS while the dependent variable in column 4 is the percentage markon. The coefficient estimates in column 1 indicate each of the above-mentioned variables is directly related to MPS with the t-ratios ranging from 2.1 to 3.5.[11] The assumed first-order process for the residuals indicates the residuals of this equation are positively correlated.[12] The serial correlation of residuals in the MPS equation could be due to a variety of reasons, for example, partial adjustment of markdowns or a left-out variable that changes slowly over time, etc. Because stores can and do change prices quickly in response to changes in demand conditions, it is unlikely the serial correlation of residuals is due to a sluggish adjustment of prices by the stores. Markdowns offered this year by a store should be determined by current demand conditions and not by the markdowns offered by the store in the previous year. A more plausible hypothesis is that the serial correlation of residuals is due to a left-out variable that changes slowly. A possible explanation for the serial correlation is the absence of a fashion variable that measures the changing use of colors and prints.

---

[11] The estimates in columns 1 and 4 were obtained using the Autoreg procedure in SAS.

[12] The correlation of residuals across the MPS and PMU equations was never very high, generally around .10. Hence, little efficiency gain was achieved when seemingly unrelated methods were employed.

The results in column 2 indicate the introduction of the lagged value of MPS does not have a significant effect on the coefficient of the import variable but does reduce the significance level of the coefficients for the percentage of shuttleless looms and the percentage of the population between 10- and 34-years old. The coefficient estimates of these two variables appear to be especially sensitive to the specification of the regression equation.

The serial correlation model implies that both the current and lagged value of each independent variable should enter as significant determinants of MPS. A bare bones version of the serial correlation model is tested in column 3 where the current and the lagged value of clothing imports are included as right-hand side variables along with the lagged value of MPS. The coefficient of the current clothing import variable is statistically significant at the 1 percent probability level and the coefficient of the lagged import variable has the correct (negative) sign and is significant at the 8 percent probability level (two-tail tests). The results in column 3 provide modest support for the serial correlation hypothesis and not for the partial adjustment hypothesis. The coefficient estimates in column 4 indicate the percent of clothing imports and the percent of young consumers are significant determinants of the percentage markon.

These regression results indicate the persistent increases in MPS and the percent markon since 1970 are related to the growth in imports. The growth of imports is indirectly related to the growing importance of fashion and product variety since the quality of imported merchandise has improved in quality as the price has risen over time. Since fashion merchandise is higher priced, the growing market share of imports implies a growth in the fashion content of products purchased by U.S. consumers. The role of the age composition of the population and of technical change is less clearly defined and only more qualified conclusions can be advanced.

III. Markups and Markdowns by Merchandise Group

Style and fashion changes are more frequent in some merchandise groups than in others. The theory of clearance sales predicts that the percentage markup will be larger and the frequency of markdowns will be higher in the groups with more frequent changes.[13] An earlier study by L. H. Grinstead (1932) documented higher markdowns relative to sales for women's fashion merchandise (outer garments) than for women's *standard* merchandise (undergarments and foundation garments), where fashion changed less frequently. MPS was also found to be higher for women's fashion clothing than for men's apparel. Historically, men's fashions have changed less frequently than have women's fashions. This historical evidence conforms closely with the implications of the clearance theory.

Contemporary data show that some of these long-term historical differences between merchandise groups have been disappearing in recent years even while new differences appear. Table 3 documents the changes that have occurred between 1965 and 1984 in markdowns relative to sales, percentage markup, and the dollar value of customer merchandise returns relative to sales. In each year, apparel merchandise has been classified into the following groups: 1) women's fashion clothing and accessories, for example, shoes, dresses, and coats, etc.; 2) women's standard or non-fashion items, which for the most part includes undergarments; 3) men's apparel and accessories; 4) teens' and junior clothing and accessories; and 5) infants' clothing. Merchandise sold in all other departments (except basements) has

---

[13] Another implication of the clearance theory is that markdowns relative to sales will be higher at the end of the season than at the beginning of seasons. Seasonal data show MPS peaks in July, the end of the spring–summer seasons, and in January, the end of the fall–winter seasons. This and other issues are discussed more fully in B. Peter Pashigian (1987).

TABLE 3—MEAN OF MARKDOWNS RELATIVE TO SALES, MARKONS, AND RETURNS RELATIVE TO SALES BY YEAR AND MERCHANDISE GROUP[a]

Year	All Groups Combined	Women's Fashion Apparel and Accessories	Women's Standard Apparel	Men's Apparel and Accessories	Teens' and Junior Apparel	Infants' Apparel	All Other Depts.
			Markdowns Relative to Sales				
1965	7.0	10.2	4.5	6.6	10.8	7.0	5.6
	(3.5)	(3.6)	(2.1)	(1.6)	(5.0)	(3.0)	(2.1)
	123	29	12	10	7	11	54
1977	10.2	11.8	8.2	10.8	18.2	10.3	8.01
	(5.1)	(6.0)	(4.1)	(4.0)	(2.4)	(4.5)	(3.7)
	130	22	16	15	11	18	48
1984	16.8	19.8	13.8	19.3	27.0	18.4	13.0
	(7.7)	(8.0)	(6.7)	(6.7)	(3.5)	(7.6)	6.1
	120	23	9	15	8	16	49
			Markons (Percentage Markup)				
1965	39.6	42.2	40.7	40.6	39.2	38.8	38.0
	(5.3)	(3.8)	(1.9)	(3.3)	(3.8)	(3.7)	(6.6)
	123	29	12	10	7	11	54
1977	46.0	48.9	48.2	48.6	49.1	46.2	42.3
	(5.8)	(2.8)	(1.8)	(2.1)	(1.4)	(4.4)	(7.4)
	130	22	16	15	11	18	48
1984	48.0	51.9	51.6	50.6	52.1	48.3	44.0
	(6.8)	(2.1)	(1.5)	(2.2)	(.4)	(4.2)	(8.6)
	121	23	9	15	8	16	50
			Merchandise Returns Relative to Sales				
1965	7.3	9.0	6.1	5.7	7.8	5.8	7.2
	(4.2)	(4.3)	(4.3)	(1.8)	(3.0)	(2.2)	(4.7)
	123	29	12	10	7	11	54
1977	7.6	8.1	7.7	8.8	10.2	6.0	7.2
	(3.6)	(3.9)	(3.0)	(2.0)	(2.0)	(2.0)	(4.3)
	110	22	12	7	9	18	42
1984	9.7	10.7	9.8	9.2	13.3	8.0	9.5
	(4.3)	(4.7)	(3.7)	(3.4)	(2.6)	(2.9)	(4.9)
	88	20	7	11	5	13	32

Note: For sources, see Data Appendix.

[a] Number in brackets represents the standard deviation. Number below standard deviation is sample size.

been combined into a catchall "other" class.[14]

Table 3 shows the simple mean, below which is the standard deviation and below the standard deviation is the number of categories within each group. In 1965 markdowns relative to sales were higher for the women's fashion group and for the teens' and junior group than for the other groups. The higher MPS for the women's fashion group than for the women's standard group or for the men's group reflects the greater importance of fashion for women's outer clothing than for women's undergarments or for men's clothing. In 1965, the percentage markup and the percentage of customer merchandise returns was highest for the women's fashion group.

---

[14] The "all other departments" category includes a diverse collection of merchandise ranging from bed sheets to small home appliances to major appliances. Some of this merchandise, for example, home furnishings, is subject to fashion trends and changes. Major durables are not immune to color and style changes but are less subject to fashion changes. The average MPS for the radio and audio appliances, television and major appliances departments increased from 5.7 percent in 1977 to 7.8 percent in 1984 or by 38 percent which is lower than the mean increase of 68 percent for the five apparel groups listed in Table 3. This comparison also indicates fashion changes have had a greater impact on apparel merchandise than on some other types of merchandise sold in department stores.

Customer merchandise returns would seem to be distantly related, if at all, to the question of markdowns and markups. But, customer merchandise returns (relative to sales) can be related to markdowns and markups if they serve as a proxy for or indicator of fashion within each group. What distinguishes fashion from non-fashion clothing is the need to match and coordinate fabric, color, pattern, and silhouette with other items and accessories. For many customers, this is not easily achieved at the point of purchase. It is not uncommon for customers to purchase several items, take them home, try them on to see if colors, patterns, and fabrics match properly. Often, merchandise is returned because colors or patterns do not match or coordinate. This is apt to be less true for non-fashion items. For apparel merchandise, returns relative to sales can serve as a useful proxy for the importance of fashion in a merchandise category.[15]

Some of these long-standing historical differences began to change after 1965. Between 1965 and 1984, markdowns relative to sales for the teens' and junior group have zoomed upward and this group now leads all other groups by a substantial margin. Because styles change rapidly in these markets, the increase in MPS is not at all surprising and is consistent with the implications of the clearance theory. While markdowns relative to sales for the women's fashion group are still greater than for the women's standard group, they are now on about par with the men's apparel group and, surprisingly with the infant clothing group. The near parity with men's apparel suggests that the role of fashion has been spreading and has become increasingly important in this group as well. The near parity of MPS for the women's fashion group with the infant clothing group is perhaps the biggest surprise. It is an unexpected finding and not easily explained by the clearance theory since style changes would not be expected to be as important and therefore as frequent for infant clothing.

The middle panel of Table 3 shows the percentage markons (markups) have also been increasing over this period. The teens' and junior group has had the largest percentage point increase from 1965–84. The role played by fashion in the merchandising of teens' and junior clothing appears to have grown in importance over time.

The lower panel also shows merchandise returns relative to sales has increased most for the teens' and junior group followed by large increases for the women's standard group and the men's apparel group. The larger increases in merchandise returns in these groups is not inconsistent with the growing importance of fashion and styles in what had been relatively staid merchandise groups. The larger increases in merchandise returns in these groups suggests that fashion has permeated these groups as well.

A plausible interpretation of these changes is that department stores are now selling more fashion merchandise than they were 20 years ago and in more apparel groups. This implies that even greater price uncertainty exists today than earlier times in determining which styles or colors will sell at initial prices. The increased price uncertainty has been reflected in both higher-percentage markons and higher markdowns relative to sales.

The quantitative effect of customer merchandise returns on MPS has been estimated with the combined use of three cross sections for 1965, 1977, and 1984. Dummy year and merchandise group effects as well as interaction effects of year with merchandise group have been estimated. Column 1 of Table 4 only includes women's fashion, women's standard, teens', and junior apparel and men's apparel groups with the men's apparel group serving as a reference group. All apparel *and* all non-apparel merchandise groups are included in column 2 with the

---

[15] In some groups high levels of merchandise returns will be caused by the incidence of damaged goods, for example, furniture, and in others by the absence of fitting rooms.

TABLE 4—MARKDOWNS RELATIVE TO SALES, 1965, 1977, AND 1984, APPAREL ITEMS ONLY AND ALL MERCHANDISE GROUPS
(t-statistic below coefficient estimate)

	Only Women's Teens', and Men's Apparel Groups (1)	All Merchandise Groups (2)
1. Constant	3.51	3.11
	2.7	3.8
2. Year, 1977	−.47	.09
	.2	.07
3. Year, 1984	2.40	2.46
	1.1	1.6
4. Women's Fashion	1.74	4.00
	1.3	4.2
5. (Women's Fashion, 1977)	−1.68	−.15
	.8	.11
6. (Women's Fashion, 1984)	−3.35	1.73
	1.8	1.2
7. Men's		1.57
		1.1
8. Men's 1977		2.60
		1.2
9. Men's 1984		4.78
		2.4
10. Women's Standard	−2.40	−.74
	1.6	.6
11. (Women's Standard, 1977)	−.99	.96
	.4	.5
12. (Women's Standard, 1984)	−1.62	3.48
	.7	1.6
13. Teens'	3.03	5.04
	1.7	3.1
14. (Teens', 1977)	1.13	3.97
	.5	1.8
15. (Teens', 1984)	.12	7.14
	.5	2.8
16. Infants		1.88
		1.4
17. (Infants', 1977)		1.66
		.9
18. (Infants', 1984)		3.96
		2.1
19. Merchandise Returns	.55	.34
	4.4	3.8
20. (Merchandise Returns, 1977)	.54	.25
	2.7	1.8
21. (Merchandise Returns, 1984)	.88	.43
	4.7	3.1
R^2 (adj)	.78	.64
Standard Error of Regression	3.49	4.05
N	151	321

Note: For sources, see Data Appendix.

"all other" group serving as the reference group.

The effect of fashion on MPS in the cross-section regressions should be captured through the coefficient of the variable for merchandise returns. If fashion is more important in a group, then customer merchandise return rates will be higher in that group. Under this assumption the merchandise returns variable will be a significant determinant of MPS. Over time, as merchandise return rates have increased, there will be a corresponding increase in MPS as long as merchandise returns is an effective proxy for fashion.[16]

Regression results are presented in Table 4. The constant in column 1 is significantly different from zero while the coefficients of the year dummy variables are not statistically significant. Hence, the *intercept* for men's clothing has not increased significantly over time. The coefficients for the other groups and group-year interaction dummy variables tend not to be significant except for women's fashion in 1984, women's standard in 1965, and teens' in 1965. More importantly, all three of the coefficients of the merchandise return variables are positive and significant. This means that in any year groups with higher merchandise return rates have higher-dollar markdowns relative to dollar sales. Assuming merchandise returns serve as a proxy for fashion, this finding suggests that fashion and MPS are directly related. The comparative size of the three coefficients means a 1 percentage point increase in the merchandise return rate increases MPS more in 1984 than it did in 1977 and more in 1977 than it did in 1965.

Very similar results are obtained when all merchandise groups are included (column 2). The reference group in this regression is the "other" group. The intercept is significant in 1984 but not in 1977, so there is some evidence that MPS increased by about 2.5 percentage points between 1977 and 1984 in the "other" group after remaining constant between 1965 and 1977. Assuming the fashion element is less important for products in the "other" group, this result suggests that some factors caused MPS to increase between 1977 and 1984. The estimated coefficients for the dummy variables indicate that the MPS for women's fashion and the teens' groups was significantly higher than the "other" group in 1965. By 1984, the MPS of the men's group, women's standard, teens', and infants has increased relative to the MPS of the "other" group.

The coefficients of the merchandise return variables are once again positive and significant. These results also indicate groups with higher merchandise return rates have higher-dollar markdowns relative to dollar sales. It is mildly comforting that the coefficients of the merchandise return variables are smaller when all groups are included than when only apparel groups are included. Because merchandise returns are less likely to be due to fashion reasons for items in the "other" group, the coefficients of the merchandise return variables would be expected to be lower when all groups are included in the regression. The increase over time of the effect of merchandise returns on MPS is left unexplained. This unexplained result suggests that the merchandise return measure is only an imperfect proxy for the importance of fashion. One obvious limitation of this analysis is that the effect of imports on MPS is not accounted for in these regressions.[17]

These results imply that the differences in MPS between groups are related in part to differences in merchandise return rates. In each of the three cross sections merchandise return rates were a significant determinant of MPS. A partial reason for the rise in MPS for teens' between 1965 and 1984 is that the merchandise return rates increased most in this group. Similarly, the narrowing of the differentials between the MPS of the women's fashion group and the MPS of the other apparel groups is that return rates have increased relatively more in these other apparel groups. These results provide some support for the notion that fashion and product diversity have grown in importance in these other apparel groups as well.

---

[16] While the rise in merchandise return rates over time could be due to the increase in lower-quality imports, this is not plausible given the rising relative price of apparel imports over time. The quality of imports has been rising, not falling over time.

[17] The absence of detailed import data for all products sold by department stores prevents a more comprehensive study.

IV. Summary

The recent and large increases in markdowns relative to dollar sales and the percentage markup appear to be related to the growing importance of fashion and product assortment. The theory of clearance sales is helpful in explaining and understanding these recent trends. It predicts a rise in the percentage markup and the dollar markdowns relative to dollar sales when price uncertainty increases and price uncertainty increases when more fashion-type merchandise is sold. Though markdowns relative to dollar sales have been historically larger for women's fashion merchandise than for women's undergarments, or for men's apparel, these differences have been narrowing since the mid-1960s. Markdowns relative to dollar sales for men's clothing are approaching those for women's fashion clothing. It is now larger for teens' and junior clothing than for any other apparel groups. This market evidence certainly suggests the role of fashion has expanded in recent years to encompass other apparel groups in addition to women's fashion clothing.

The clearance theory has proven helpful in understanding recent time-series trends, in explaining differences in markdown policies between merchandise groups and in explaining why markdowns peak in July and January, the end of the spring–summer and fall–winter seasons. These major regularities are not easily explained by other theories of sales. Still, not all sales can be explained by the clearance theory. Temporary promotional sales cannot be avoided by even the most casual reader of newspaper ads. These short-term price promotions are more difficult to explain. It is unclear whether existing primitive theories of temporary sales have enough testable implications to explain this type of sale.

The question of why fashion has become more important in recent years has been touched upon but only briefly. Fashion is undoubtedly a superior good. Yet markdowns and markups only began to rise around 1970 even though real per capita income has been increasing throughout the post-World War II period. This suggests the underlying explanation for the growing importance of fashion apparel may be found on the supply side and not the demand side. Though the preliminary time-series results failed to offer confirming evidence, the trend toward fashion may be due to the recent technological changes in the weaving and knitting of fabric. These changes have lowered the set-up cost of changing designs and patterns and thereby reduced the cost of producing short runs of fashion fabric. The deeper cause for the growing role of fashion may spring from these technological changes.

DATA APPENDIX

Figure 1:

1925–1960: Markdowns relative to dollar sales and markon reported in Malcolm P. McNair and Eleanor G. May, *The American Department Store, 1920–1960*, pp. 22–25.

1961–1965: Markon estimated by taking the markon figures reported in National Retail Merchants Association, *Merchandising and Operating Results of Department and Specialty Stores in 1965* and multiplying by 1.054 to adjust for the average percentage difference (in 1959 and 1960) between the level of markon reported by McNair and May and by the National Retail Merchants Association.

1966–1984: Markdowns relative to dollar sales and markon reported in selected issues of National Retail Merchants Association, *Department and Specialty Store Financial and Operating Results*, annual.

Figure 2:

Data for color of sheets are reported in Bureau of Census, *Sheets, Pillowcases, and Towels*, Current Industrial Reports, MQ 23X. Market shares are based on value of shipments of flat and fitted sheets (except crib) by U.S. mills. Market shares are calculated from the sale of all types of sheets from 1960–74; from the sale of polyester-cotton blends from 1975–82 (major share of the market); and from all sheet sales from 1983–86.

Figure 3:

Men's Dress Shirts: Data for color of men's dress shirts were kindly supplied by Cluett Peabody and Company and are based on diary records of a panel sponsored by the *Market Corporation of America*. These data are purchased and used by numerous retailing and the textile firms to study market trends. Panel composition has changed periodically over the years so the data are not strictly comparable from year-to-year (the latest change occurred in 1983). Still, the large change from the 1960s to the 1980s in color used is primarily due to changes in the demand for colors and not to changes in the composition of the panel.

Women's Sportswear: Data for the market share of women's sportswear are derived from data published in the *Census of Business-Retail Trade*, Merchandise Line Sales.

For 1963–1967, the total market is defined as the sum of dollar sales of dresses (merchandise line code MLC 167 for department stores and MLC 172 for women's ready-to-wear stores) and women's sportswear (MLC 168 for department stores and MLC 168 for women's ready-to-wear) and is limited to combined sales of dresses and sportswear reported by department stores (SIC 531) and women's ready-to-wear (SIC 562) stores.

For 1972, dollar sales of dresses are for MLC 167 for department stores, MLC 224 for women's ready-to-wear, and family clothing stores (SIC 565). Dollar sales of women's sportswear are for MLC 226 for department stores, ready-to-wear, and family clothing stores.

For 1977, dollar sales of dresses are for MLC 224 for department, ready-to-wear, and family clothing stores. Dollar sales of sportswear are for MLC 227 and 228 by department, ready-to-wear, and family clothing stores.

For 1984, dollar sales of dresses are for MLC 223 by department, ready-to-wear, and family clothing stores. Dollar sales of sportswear are the sum of sales totals for MLC 225, 226, 227, and 228. Suits are combined with pantsuits, sports jackets, and blazers in MLC 225 and cannot be excluded from the sportswear totals in 1982. Suits are excluded from the sportswear totals in previous years.

Table 2:

Clothing Imports Divided by Shipments of Apparel Industry:

Clothing imports reported in annual issues of *Statistical Abstract of the United States.* Figures originally reported in U.S. Department of Commerce, International Trade Administration, Overseas Business Reports, *U.S. Foreign Trade*, annual.

Value of Shipments for the Apparel and Other Textile Products (SIC 23), U.S. Bureau of Census, *Annual Survey of Manufacturers*, annual. Shipments from 1956 to 1960 are not available. Shipments for these years were estimated by multiplying the reported industry value added totals in each year by 2.214725. Percent of the Population between 15 and 34: U.S. Bureau of Census, *Current Population Reports*, Series P-25, Nos. 985,917,721,519,311.

Percent of Shuttleless Looms: U.S. Bureau of Census, *Broadwoven Fabrics* (Gray), Current Industrial Reports, MQ 22T. Number of shuttleless looms first reported in 1968. Market share of shuttleless looms in 1968 was 2.2 percent. Market share of shuttleless looms is assumed to be zero from 1956 to 1965. Market share of shuttleless looms was linearly extrapolated between 1965 and 1968.

Tables 3 and 4:

Data are taken from annual publication by the National Retail Merchants Association in 1965, 1977, and 1984. The 1965 publication is entitled, *Merchandising and Operating Results of Department and Specialty Stores in 1965.* The 1977 publication is entitled, *Department and Specialty Store Merchandising and Operating Results of 1977.* The 1984 edition is entitled, *Department and Specialty Store Merchandising and Operating Results of 1984.* These publications report operating results by department or demand center. The classification system for departments was revised between 1965 and 1977 so the system used in 1965 is different from the systems used in 1977 and 1984. The definition of each of the variables is reported in each publication. The author classified the departments into the groups reported in Tables 3 and 4.

REFERENCES

Arpan, Jeffrey S., de la Torre, Jose and Toyne, Brian et al., *The U.S. Apparel Industry: International Challenge, Domestic Response*, Georgia State University, 1978.

Cline, William R., *The Future of World Trade in Textiles and Apparel*, Institute for International Economics, 1987.

Greenwood, Kathryn M. and Murphy, Mary F., *Fashion Innovation and Marketing*, New York: Macmillan, 1978.

Grinstead, L. H., "The Reasons for the Extent of Markdowns in Ohio Department Stores," Bureau of Business Research, College of Commerce and Administration, Ohio State University, 1932.

Lazear, Edward P., "Retail Pricing and Clearance Sales," *American Economic Review*, March 1986, *76*, 14–32.

McNair, Malcolm P. and May, Eleanor G., *The American Department Store, 1920–1960*, Bureau of Business Research, Bulletin No. 166, Harvard University, 1963.

Olsen, Richard Paul, *The Textile Industry*, Lexington: Lexington Books, 1970.

Pashigian, B. Peter, "Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing," *Center for the Study of the Economy and the State*, Working Paper No. 49, 1987.

Salop, Steven and Stiglitz, Joe E., "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents," *American Economic Review*, December 1982, *72*, 1121–30.

Varian, Hal R., "A Model of Sales," *American*

Economic Review, September 1980, *70*, 651–59.

Financial Executives Division, National Retail Merchandise Association, Merchandising and Operating Results of Department and Specialty Stores, New York, New York.

U.S. Bureau of Census, *Annual Survey of Manufacturers*, annual.

______, *Census of Business-Retail Trade*.

______, Current Industrial Reports, *Broadwoven Goods*, (MQ-22T).

______, Current Industrial Reports, *Sheets, Pillowcases and Towels* (MQ23X).

______, *Current Population Reports*, Series P-25.

______, *Statistical Abstract of the United States*, Washington: USGPO, annual.

U.S. International Trade Commission, *U.S. Global Competitiveness: The U.S. Textile Mill Industry*, USITC Publication 2048, December 1987.

Franchising and Risk Management

By ROBERT E. MARTIN*

Franchising is an important and controversial form of vertical integration. Allegations of opportunistic behavior by franchisors have led to calls for public regulation and in some states "fairness in franchising" laws. The advisability of such regulation depends on the long-run incentives to franchise. This paper offers new evidence on the incentives to franchise.

Why do some firms rely on expansion by franchising rather than expansion by company-owned outlets? Why do many more firms rely on a mix of franchising and company-owned outlets for expansion? There are two traditional explanations offered by the literature. Many practitioners and some economists argue that franchising is a means of raising scarce capital.[1] Due to the capital market imperfections implied by this argument, it has not been well-received.[2] Alternatively, franchising may be a response to the monitoring problems encountered in remote and dispersed locations. The existence of parallel forms of vertical integration (company-owned operations with franchised operations) can be explained by variations in the cost of monitoring.[3] The cost of monitoring a single site in a remote location may be higher than the average cost of monitoring numerous outlets in a dense urban area.

Another related set of incentives to franchise are identified and tested in this paper. The exploitation of economies of scale in promotion and in monitoring requires large numbers of outlets at heterogeneous locations. Different locations have different expected returns and different "risk" characteristics. Franchising allows the firm to exploit these economies of scale and at the same time to shed "risky" locations and retain more profitable sites as company-owned outlets. The franchisor may choose to franchise risky locations as a consequence of simple risk aversion or, if the franchisor is risk neutral, because monitoring costs rise as risk increases.[4] A location with substantial variation in sales will require closer supervision as a company-owned outlet than a location with very little variation in sales. Hence, higher risk may induce higher monitoring costs. The franchisor's ability to shift risk to franchisees will be limited by the franchisee's risk aversion[5] and by the expected profitability of the sites. Franchisees will be will-

---

*Department of Economics, Louisiana State University, Baton Rouge, LA 70803. Many useful comments were received from members of the Workshop Series at LSU. A very special debt is owed to an anonymous referee, Carter Hill, Bob Newman, and Lee Adkins for their comments and suggestions.

[1] For a more complete discussion of the nature of franchising incentives, see Richard Caves and William Murphy, 1976; and Frank Mathewson and Ralph Winter, 1985.

[2] Paul Rubin, 1978, argues that the capital shortage incentive to franchise is in conflict with modern capital theory. He maintains that selling franchise shares to franchisees is less costly. Hence, raising capital by selling franchises would be inefficient.

[3] James Brickley and Frederick Dark, 1987, provide empirical evidence that suggests a direct relationship between monitoring costs and franchising.

[4] I am indebted to an unknown referee for observing that "risk" may influence the franchisor's organizational mix through monitoring costs even if the franchisor is risk neutral.

[5] It has been observed by Rubin (1978, pp. 225–26) and others that the representative franchisee's investment is undiversified relative to the representative franchisor. From this it is concluded that the franchisee will be more risk averse than the franchisor. I suggest that the *choice* of a less diversified investment reflects *less* risk aversion on the part of franchisees. Recall that franchisees are drawn from the more entrepreneurial agents in our society, since they chose to start their own business rather than seek employment.

ing to accept higher risk if they are rewarded appropriately with higher expected returns.

The paper proceeds as follows. Section I contains a discussion of the franchising choice under risk and variable monitoring costs. Section II contains a long-run empirical model using firm-specific data. Franchising life-cycle functions are estimated in this section. The capital shortage, relative profitability, and risk hypotheses are tested in the short-run model in Section III. The final section contains a summary of the implications.

I. Monitoring and Risk in Franchising

Agency problems abound in the choice of organizational form.[6] The nature of most franchised businesses is such that there are numerous, geographically dispersed production centers. Geographic dispersion creates special monitoring problems for the firm. Managers of company-owned outlets may shirk since their supervisors are not in permanent residence. Greater dispersion in locations implies more supervisors and higher monitoring cost since more time is lost moving between locations. Shirking incentives are reduced by the franchise agreement since the franchisee has a claim on the residual. Hence, we may expect that remote locations will be franchised and that geographically concentrated locations will be retained as company outlets. James Brickley and Frederic Dark's (1987) empirical evidence suggests that franchised units are more likely to be located in rural and small urban areas while company-owned outlets are concentrated in larger urban areas.

Franchised outlets must also be monitored and the franchisors incur opportunity costs for each site that is franchised. The franchisor must forego income equal to the franchisee's profit share net of the earnings from the reinvested franchise fee. Franchisees have an incentive to free ride against the quality choice of other franchisees and the company-owned outlets. In order to preserve the brand-name capital, which is the principal asset in many cases, the franchisor must monitor the franchisee's quality choice. The need to monitor the franchisee's quality choice declines as the proportion of repeat customers increases. The market will monitor the quality choice when transient customers are rare. The highest proportion of repeat customers is most likely to be observed in rural and small urban areas. The highest proportion of transient customers is most likely to be observed in urban areas with high population densities and along high traffic corridors. Hence, quality monitoring reinforces the tendency to franchise outlets in the more remote locations. In any event, monitoring costs are a very important element in the selection of organizational form.

There are also substantial incentives for the firm to use franchising move intensively in the short run, beyond what might be implied by monitoring costs. Richard Caves and William Murphy (1976), and Frank Mathewson and Ralph Winter (1985) argue that economies of scale exist in the promotion of national brand names. The minimum efficient scale for the exploitation of national or regional brand names would be measured by the number and dispersion of outlets (either franchised or company-owned). Little is to be gained by mass-media advertising in markets where the firm has few outlets. The gains from advertising rise as the number of outlets increases. The young firm may have a competitive imperative to reach the minimum efficient scale as quickly as possible.[7] Since franchised outlets require less monitoring, the firm can expand more rapidly with franchised outlets than with company-owned outlets. As the firm matures, franchise contracts expire, and the firm will begin to exploit economies of scale in the monitoring of company-owned outlets. Hence, the proportion of company-owned outlets as a function of years in franchising should exhibit a convex structure. We may also expect the term of franchise contracts to be shorter for young firms than for mature firms.

---

[6] Representative discussions of the agency problems encountered can be found in Brickley and Dark, 1987; Mathewson and Winter, 1985; and Rubin, 1978.

[7] Patrick McGuire, (1971, p. 6), reports that "...the single most important reason for adopting franchised distribution would be to conserve or acquire capital, while at the same time attempting to establish an effective distribution network as quickly as possible."

Heterogeneous locations also imply differences in expected profitability and risk by location. If the franchisor is risk neutral or risk averse, that person will retain locations with high expected profitability since the opportunity cost of franchising is higher for these locations. The firm's cost of franchising rises relative to the cost of monitoring company outlets as expected profitability increases. Hence, there is a strong incentive to retain more profitable sites as company-owned outlets.

The role of risk in the choice of organizational form is more debatable, however. It may be argued that firms are risk neutral, since stockholders can costlessly diversify their portfolios.[8] Observed behavior and the extensive theory of the firm under uncertainty suggest that firms have nonlinear attitudes toward risk. For example, the risk-neutral firm would not buy insurance.[9] Yet, firms spend more for insurance premiums than they pay in dividends.[10] Louis DeAlessi (1987) notes that the controlling stockholders of many firms do not completely diversify their portfolios, even though the transactions cost of such diversification are negligible. He cogently argues that incomplete diversification by some shareholders is optimal. There is a tradeoff between diversification and control. Complete diversification by all stockholders offers little incentive to monitor the firm's management and such monitoring is not a costless endeavor. Hence, partial diversification by some stockholders is in the interest of all stockholders. The undiversified portfolios of the controlling stockholders imply that risk does matter.

Risk may also affect the cost of monitoring company-owned outlets. Greater variability in sales implies more frequent periods when sales are less than expected. Periods with low sales require more management inputs and this implies rising monitoring costs. If monitoring costs rise with variability, "risky" locations will be franchised. Variability would lead to franchising in this case, even if franchisors are risk neutral.

There is substantial observational equivalency between the monitoring incentive to franchise and the risk avoidance incentive to franchise. High population densities and diversified local economies in urban areas should have a tendency to smooth sales. Hence, risk should be lower in urban areas than in remote rural areas with undiversified local economies. Risk avoidance would imply that urban locations will be retained as company-owned outlets, just as lower monitoring costs would imply that these locations will be retained by the franchisor.

Some insights into these issues can be obtained by comparing the performance of the representative company-owned outlet with the performance of the representative franchised outlet by industry. The data for the performance ratios contained in Table 1 cover the period from 1969 through 1986. The average sales ratio in column 1 is obtained by dividing average sales per company-owned outlet by average sales per franchised outlet. Note that this ratio is greater than one for every industry except one and that the average value for all industries is over three. These statistics are consistent with the existence of scale economies in monitoring and the hypothesis that franchisors retain locations with the highest expected profits as company-owned outlets. The statistics in columns 2 and 3 are obtained by estimating an exponential sales growth func-

[8] If the market for managers is imperfect and a substantial portion of a management's human capital is in firm-specific investments, the firm may be risk averse. De Alessi's, 1987, explanation of why firms have nonlinear attitudes toward risk does not depend on imperfect markets.

[9] Since insurance premiums are not actuarially neutral, the purchase of insurance would reduce stockholders' equity. Some have argued that firms hire the insurance company's expertise and it is not insurance that they purchase. This argument is weak. The expertise can be hired without also purchasing insurance. Why would the risk-neutral firm purchase the bundled product when they are available separately?

[10] David Mayers and Clifford Smith (1982, p. 281), report that business firms paid some $12 billion more in insurance premiums than they paid in dividends in 1978.

TABLE 1—PERFORMANCE RATIOS[a] 1969–1986

Industry	Average Sales per Outlet	Compound Growth Rate in Sales per Outlet	Standard Regression Error[b]
Automotive Products and Services	3.63	1.45	1.70
Accounting, Credit, Collection, and Business Services	4.42	1.16	.84
Employment Services	1.70	.90	.16
Printing and Copying Services	1.10	.98	.72
Miscellaneous Business Services	5.23	1.92	1.14
Const., Home Imp., Maint. and Cleaning Services	7.81	2.14	4.82
Convenience Stores	.97	1.26	1.91
Educational Products and Services	1.39	1.74	.39
Restaurants	1.28	.78	.28
Hotels, Motels, and Campgrounds	2.10	1.48	.42
Recreation, Ent., and Travel	6.81	1.43	4.15
Rental Services	5.12	.63	.35
Retailing	2.26	.34	6.16
Gasoline Service Station	1.02	.98	.74
Laundry, Drycleaning Service	2.69	2.51	2.51
Miscellaneous	1.96	.62	1.34
Average	3.11	1.27	1.73

Data Source: Franchising in the Economy.

[a] The relevant company-owned statistic divided by the franchised statistic.

[b] $[\sigma c/\sigma f]$ where σ is the standard error of the regression. In each case, the natural log of sales per outlet was the dependent variable and time was the independent variable.

tion for both company-owned sales per outlet and franchised sales per outlet. Column 2 reveals that company-owned sales per outlet grew more rapidly than franchised sales per outlet over the same period in 9 of the 16 industries and that the average ratio was greater than one.

The statistic in column 3 of Table 1 is a measure of the "risk" in company-owned sales relative to the "risk" in franchised sales per outlet.[11] The standard regression errors from the respective exponential sales growth functions were used to calculate the ratio of the standard regression errors. The regression error measures the unexplained variations in sales from the estimated exponential growth function. The ratio in column 3 is less than one in 8 of the 16 industries, while the average is greater than one. The mixed results across industries regarding relative risk bearing may reflect reputation effects, as well as relative risk aversion between franchisors and franchisees. Mathewson and Winter (1985, p. 503) report that many franchisors face persistent queues of potential franchisees and the long-term growth prospects vary substantially across industries.[12] The franchisors in the favored industries would be in a position to offer less advantageous terms to franchisees than franchisors in less favored industries.

---

[11] Again, I am indebted to an unknown referee for suggesting this risk measure.

[12] The variation in growth rates by industry is readily revealed by the growth functions estimated for sales per outlet.

II. Franchising Life Cycles

Franchising life-cycle functions for each industry are estimated in this section. These functions have two applications: They reveal the representative firm's long-run proportion of company-owned outlets (PCO) and they are used to control for the effect of franchise system maturity in the industry model in Section III. There are 949 franchising organizations represented in the firm-specific sample. The firms are distributed among 16 different industries. The data for each organization consist of the number of years the firm has been franchising (YF), the number of company-owned outlets (COE), and the number of franchised outlets (FOE). All of these data were obtained from *The 1985 Franchise Annual* and the *Source Book of Franchise Opportunities*.

Consider the implications of the extreme forms of the capital shortage hypothesis and the monitoring hypothesis. If capital shortages are the only reason for franchising, then franchising will be a transitory phenomenon. As the firm's liquidity improves with maturity, PCO will converge to one. If franchising is a superior incentive system and monitoring costs are uniform, then company-owned outlets will be a transitory phenomenon. As franchised outlets drive alternative forms of organization out of the market, PCO will converge to zero. The strict long-run versions of these two hypotheses have sharply contrasting implications. Thus, an examination of the relationship between the life cycle of the firm, as measured by years in franchising (YF), and the proportion of company-owned outlets is in order.

A. *Life-Cycle Functions*

The estimation of PCO life-cycle functions is a limited dependent variable problem.[13] PCO is bounded on the unit interval. If YF = 0, then PCO = 1. The logit transformation ensures that the predicted values for PCO will be restricted to the unti interval.[14] The loglinear function to be estimated is

$$\text{(1)}\quad \ln[\mathrm{PCO}_i/(1-\mathrm{PCO}_i)] = \theta_1\cdot \mathrm{YF}_i + \theta_2 \mathrm{YF}_i^2 + \theta_3\cdot \mathrm{YF}_i^3 + e_i/[P_i(1-P_i)],$$

where P_i is the true population value of PCO_i. Equation (1) is estimated with weighted least squares, where $\mathrm{PCO}_i(1-\mathrm{PCO}_i)$ is a consistent estimator of $P_i(1-P_i)$. A more complete discussion of the estimation technique can be found in George Judge, William Griffiths, Carter Hill, and Tsoung-Chao Lee (1980, pp. 592–93).

The estimated logit life-cycle equations are contained in the Appendix. Fourteen of the sixteen equations are significant at the .10 or better level. Plots of the life-cycle functions over a 30-year period reveal that five of the functions decline monotonically, seven of the functions are convex with unique interior minimums, three of the functions have both an interior minimum and maximum, and one is concave with a unique maximum. The three functions with both an interior minimum and maximum are convex over the first 15 years in franchising. The concave function is one of the two logit functions that was not significant. The other insignificant function was one of the monotonic functions. Thus, 10 of the 14 significant equations exhibit a convex structure. These results are consistent with the existence of economies of scale in the promotion of national brand names and the competitive imperative to reach the minimum efficient scale as soon as possible.

Table 2 contains three separate estimates of the long-run PCO for each industry. The life-cycle estimate of long-run PCO is obtained by computing the logit estimate for the year in franchising that lies one standard deviation above the average number of years

---

[13] A general discussion of the empirical problems encountered in this case can be found in Judge, Griffiths, Hill, and Lee (1980, pp. 582–621).

[14] See the weighted least squares model of the logit estimator in Judge, Griffiths, Hill, and Lee (1980, pp. 592–93).

TABLE 2—LONG-RUN PROPORTION OF COMPANY-OWNED OUTLETS: LOGIT LIFE-CYCLE FUNCTIONS

	Industry Average	Life Cycle		Oldest Firm		Mature Firms		Young Firms	
Industry	PCO	Year[a]	PCO	Years[b]	PCO	Years[b]	PCO	Years[b]	PCO
Automotive Products Services	.20	23	.05	50	.09	24	.14	7	.22
Accounting, Credit, Collection, and Business Services	.13	22	.24	39	.00	21	.08	6	.16
Employment Services	.21	19	.34	33	.15	19	.21	5	.21
Printing and Copying Services	.13	15	.02	22	.39	16	.06	5	.17
Miscellaneous Business Services	.15	12	.40	24	.33	12	.11	4	.18
Const., Home Imp., Maint. and Cleaning Services	.16	16	.11	40	.00	15	.10	5	.20
Convenience Stores	.22	23	.43	56	.00	23	.09	5	.30
Educational Products and Services	.27	15	.21	25	.04	15	.22	4	.30
Restaurants	.32	20	.36	59	.02	21	.25	5	.36
Hotels, Motels, and Campgrounds	.18	41	.28	73	.56	44	.27	14	.15
Recreation, Ent., and Travel	.13	19	.04	31	.02	19	.04	4	.19
Rental Services	.16	26	.38	60	.28	27	.17	6	.15
Retailing	.16	18	.55	56	.00	19	.25	4	.31
Real Estate	.15	19	.38	40	.00	21	.11	6	.17
Consumer Services	.26	15	.29	49	.03	14	.19	4	.30
Miscellaneous	.23	25	.70	72	.09	26	.09	3	.28

[a] Industry average YF plus one standard deviation.
[b] Number of years in franchising.

in franchising for each industry. The evaluation of the logit life-cycle functions at one standard deviation above the industry average is admittedly arbitrary. The purpose is to obtain a point estimate that reflects the PCO of a mature firm, one that is well within the sample experience and not from the upper tail of the age distribution. The second estimate of the long-run PCO by industry is the actual proportion of company-owned outlets for the oldest firm in each industry. The third estimate of the long-run PCO by industry is the average PCO for the mature firms in each industry. The mature firms are defined to be those firms that have been in franchising for a period longer than the average number of years in franchising for each industry. The last column in Table 2 contains the average PCO for the young firms, those firms who have been in franchising for a period less than the average number of years in franchising.

A review of the results in Table 2 reveals that none of the proportions are within 10 percentage points of one. Only three of the proportions are greater than .5 and in each of these three cases the alternative estimates are substantially lower. The highest value was .70 for the logit estimate in the miscellaneous category. The other two estimates in the miscellaneous category imply that the long-run PCO is approximately .09. Alternatively, 13 of the 48 PCOs estimated in Table 2 are between zero and .05. Consistently small estimates are found in one of the 16 industries and 9 of the 16 oldest firms have PCOs between zero and .05. Twenty-nine of the 48 PCOs are between .05 and .50. The 95 percent confidence intervals for the life-cycle point estimates are contained in the Appendix. The upper bounds are all significantly less than one, except for the miscellaneous category. Four of the lower bounds include zero. The intervals are consistent with the hypothesis that long-run PCOs are bounded away from one, but not away from zero. Franchising is a permanent form of distribution in each industry and the long-run values appear to lie between zero and .50.

B. *Profitability*, *Risk*, *and the Long-Run PCO*

The life-cycle functions allows us to control for the age of the representative franchisor. In principal, it should be possibly to directly test whether the long-run PCO depends on relative profitability and relative risk. For example, we might include the ratio of average sales (RAS_i), the standard error from the loglinear function estimated for company-owned sales per outlet (SEC_i) and the standard error from the loglinear function estimated for franchisee-owned sales per outlet (SEF_i), in an aggregate logit life-cycle function which also includes the number of years in franchising as a third-degree polynominal. There are estimation problems with this approach, however. The long-run PCO is chosen from a portfolio of locations with different expected profitability and risk. In turn, the short-run PCOs determine RAS_i, SEC_i, and SEF_i. Therefore, long-run PCO_i, RAS_i, SEC_i, and SEF_i are simultaneously determined. The estimation of equation (3) with RAS_i, SEC_i, and SEF_i will suffer from simultaneity bias.

A solution is found by using the logit life-cycle functions as instrumental variables for the estimation of the long-run PCO in a two-step procedure. In each industry, $\overline{PCO}_i$ is the predicted long-run PCO for the year in franchising that lies one standard deviation above the average YF for that industry. The equation estimated with weighted least squares is

$$(2)\quad \overline{PCO} = \underset{(8.355)}{1.1946} - \underset{(2.829)}{.0451}RAS + \underset{(1.576)}{.0033}SEC - \underset{(3.005)}{.0053}SEF$$

$$\bar{R}^2 = .5740 \quad F = 6.840,$$

where the absolute t values are in parentheses. The signs and the significance levels are qualitatively the same as those we obtain by estimating the logit function in equation (1) with RAS, SEC, and SEF. Equation (2) reveals that low long-run PCOs are associated with *higher* levels of company-owned sales per outlet relative to franchised sales per outlet. Low PCOs are also associated with *less* "risk" in company-owned outlets and with *more* "risk" in franchised outlets. The more selective the franchisor is about the outlets that person retains as company-owned units, the more profitable and less risky they appear relative to the franchised units. The representative franchisor's ability to be selective in the retention of company-owned outlets most probably depends on differences in reputation effects across industries.

III. Industry Model

Considerable variation can be observed in the short-run proportion of company-owned outlets. The proportion varies across industries and over time. The purpose of this section is to test whether credit market conditions, the life-cycle functions, relative profitability, and risk influence the short-run variations in PCO. The short-run model provides an opportunity to discriminate between the risk avoidance effect and the risk-induced higher monitoring cost effect.

The short-run proportion of company-owned outlets for the ith industry in the tth time period is PCO_{it}. PCO_{it} is equal to the total number of company-owned outlets in the industry divided by the sum of both types of outlets. Hence, PCO_{it} is the proportion chosen by the representative firm in the ith industry during the tth time period. Company-owned sales per outlet and franchised sales per outlet are CSO_{it} and FSO_{it}, respectively. PCO_{it}, CSO_{it}, and FSO_{it} provide the foundation for the empirical model of the representative franchisor by industry. The industry data consist of time-series and cross-sectional information for eighteen franchising industries. The time period covers 1969 through 1986. These data were collected from *Franchising in the Economy*.

A. *The Short-Run Model*

Tight credit conditions imply high real rates of interest and in some instances credit rationing. In any event, marginal borrowers will be rationed out of the market by either the high real interest rates or the loan qualification process. If the franchisor is unwilling to adjust the mix of company owned to franchised outlets at the margin, that person must forego attractive opportunities and experience a slower rate of growth. Suppose the franchisor is a mature firm with established brand-name capital. Tight credit conditions may reduce the number of qualified franchisees relative to the number of attractive sites this franchisor has available. If the franchisor does not retain these sites as company-owned outlets, they may be lost or, worse yet, lost to a competitor. If the established franchisor retains the surplus sites as company-owned outlets, that person also retains the opportunity to sell them as franchised outlets at a later date when credit conditions improve.

If the franchisor is a young immature firm, that person may be the marginal borrower relative to qualified franchisees. Or, the terms on which the firm can borrow may represent an unacceptable dilution of equity.[15] The failure to adjust its PCO at the margin would represent the loss of attractive sites and a slower rate of growth. If there is a competitive imperative to reach the minimum efficient scale as soon as possible, an additional incentive to franchise sites the firm would otherwise retain as company-owned outlets would exist. The franchisor can recapture the site as a company-owned outlet when the franchise contract expires. Although both mature and immature franchisors may adjust short-run PCO in response to credit market conditions, the adjustments are in the opposite direction. We expect to find the mature firm's PCO rising during tight credit conditions and the immature firm's PCO declining during tight credit conditions. These characteristics should be reflected across industries, in part, by differences in total investment required per outlet and the ratio of mature to immature franchisors. A real interest rate variable (RINT_{t-1}) is included in the regressions in order to control for credit market conditions. RINT_{t-1} is the difference between the federal funds rate and the annual percent change in the GNP implicit price deflator in $t-1$. A separate real interest rate coefficient is estimated for each industry.

Risk aversion and risk neutrality imply that the franchisor will retain the sites with higher expected profits as company-owned outlets. Subjective profit expectations imply that the franchisor will alter the short-run PCO at the margin if observed relative profitability deviates substantially from expected relative profitability. Since the franchisor does not have perfect foresight with respect to profits *or* the moments of the true distribution of random profits, one can expect to observe learning behavior in the short run.

Suppose the relationship between profit margins on company-owned sales and franchised sales remains stable over the cycle. Then, the ratio of company-owned sales per outlet to franchised sales per outlet, say RS_{it}, will be an accurate measure of relative profitability. The expected ratio of sales is the out-of-sample average RS for the preceding five years. Thus, ERS_{it} is the average value for RS from $t-5$ through $t-1$. The unanticipated deviation in relative profitability observed in $t-1$ is $\mathrm{UDRS}_{it-1}=\mathrm{RS}_{it-1}-\mathrm{ERS}_{it-1}$. If observed relative profitability is equal to expected relative profitability, the franchisor would have no incentive to alter the short-run PCO in time period t. The first five years of observations for each industry are sacrificed in order to calculate UDRS. Another year is lost in the lag operation, since PCO_{it} depends on UDRS_{it-1}. More importantly, the out-of-sample techniques

---

[15] By definition, a young or immature franchisor is a firm that is unlisted and does not have access to national capital markets. Funds for expansion by company-owned outlets must be obtained from venture capitalists, who will require a substantial portion of the equity in exchange. Venture capital funding may involve loss of the controlling interest. The rewards for entrepreneurship are the capital gains obtained from "taking the organization public." Hence, dilution of the equity is the franchisor's primary concern. Franchising allows the firm to achieve "off balance sheet" financing that does not dilute equity.

TABLE 3—FIRST EQUATION,[a] PROPORTION OF COMPANY-OWNED OUTLETS

		$F = 1222.55$	$\bar{R}^2 = .9957$		
Industry	Intercept Terms β_o^i	Interest Rates Ω^i	Profitability β_1	Company-Owned Risk β_2	Franchised Risk β_3
			.00216 (3.072)	−.04952 (3.206)	.02585 (1.020)
Automotive Products and Services	.10255 (33.308)	.00393 (4.323)			
Accounting, Credit, Collection Agencies, and General Business Services	.02463 (3.848)	−.00151 (1.001)			
Employment Services	.26985 (19.317)	.01453 (3.478)			
Printing and Copying Services	.05473 (5.442)	−.00598 (2.512)			
Tax Preparation Services	−.00718 (46.585)	.48414 (2.317)			
Misc. Business Services	.03508 (4.756)	.00090 (0.627)			
Construction, Home Improvement, Maint., and Cleaning Services	.04588 (8.057)	−.00238 (1.774)			
Convenience Stores	.64543 (83.946)	−.01359 (5.932)			
Educational Products and Services	.18436 (10.920)	−.01850 (4.687)			
Restaurants	.28774 (76.266)	.00722 (6.702)			
Gasoline Service Stations	.19120 (60.420)	−.00237 (3.407)			
Hotels, Motels, and Campgrounds	.15771 (16.926)	−.00230 (0.903)			
Laundry and Drycleaning Services	.02307 (3.795)	.00294 (1.420)			
Recreation, Entertainment and Travel	.02765 (5.248)	.00203 (1.646)			
Rental Services	.23744 (26.331)	−.00749 (3.158)			
Retailing	.28143 (59.649)	.00702 (5.361)			
Soft Drink Bottlers	.04330 (5.512)	.00388 (1.374)			
Miscellaneous	.09535 (8.435)	.00046 (0.156)			

[a] All absolute t values are in parentheses.

used to calculate $UDRS_{it-1}$ avoid the simultaneity problem identified in the previous section.[16]

The same measure of risk will be employed in the short-run model as in the long-run model, except that the standard errors are obtained from out-of-sample estimations. SRC_{it} is the sales risk variable for company-owned outlets and SRF_{it} is the sales risk variable for franchised outlets. SRC_{it} is equal to the standard error of the company-owned sales regression model over time period $t-4$ through t divided by the predicted value for t. The independent variable in the sales regressions was time. Two model specifications were estimated: A simple linear functional form and an exponential functional form. After forecasting one year in advance, the root mean square error (RMSE) of these forecasts was calculated for the 12-year period. The functional form with the lowest RMSE was used to generate SRC_{it} and SRF_{it}. The linear functional form proved to have the lowest RMSE in most cases.

---

[16] For example, regressing sales for each industry against time over the entire data set, computation of the predicted value for each year, and then taking the ratio of sales would introduce simultaneity bias.

The first equation estimated was

$$\text{(3)}\quad \mathrm{PCO}_{it} = \sum_{i=1}^{18} \beta_0^i + \beta_1 \mathrm{UDRS}_{it-1} + \beta_2 \mathrm{SRC}_{it-1} + \beta_3 \mathrm{SRF}_{it-1} + \sum_{i=1}^{18} \Omega^i \mathrm{RINT}_{t-1},$$

where the β_0^i's are intercept terms for each industry and the Ω^i's are industry coefficients for the effect of real interest rates. Weighted least squares was employed in the estimation of all equations. The relative profitability hypothesis implies that β_1 should be positive. The risk-avoidance effect and the risk-induced monitoring cost effect imply that β_2 should be negative, when we control for franchise risk with SRF and relative profitability with UDRS. The greater the unexplained proportional variance in company-owned sales per outlet in the previous period, the lower PCO should be in the current period.

The results from the estimation of the first equation are contained in Table 3. Note that β_1 is positive and significant at better than the .01 level, while β_2 is negative and significant at better than the .01 level. The coefficient for SRF is positive, but not significant. The risk in company-owned sales per outlet is significant in explaining short-run variations in PCO, while the risk in franchisee's sales per outlet is not significant. All of the intercept terms are positive and significant at better than the .01 level. Nine of the interest rate coefficients are negative, and seven of these are significant at the .15 or better level. Nine of the interest rate coefficients are positive and five of those are significant at the .15 or better level. A joint F test on all of the real interest rate coefficients was conducted. The test value was $F=12.54$, which is significant at better than the .01 level. The F value for the entire equation was $F=1266.44$. These results are consistent with the relative profitability hypothesis, the risk effects, and the short-run capital market effects.

The life-cycle models in Section II imply that the number of years in franchising will influence the proportion of company-owned outlets. The industry definitions in the firm data set and in the industry data set are consistent for 14 of the 18 industries. Thus, we have prior information about the industry life cycle for 14 of the 18 industries. We cannot say what the effect of the number of years in franchising is on PCO for the remaining four industries, so we exclude them from this portion of the analysis.[17] A pro forma proportion of company-owned outlets, say PCOF_{it}, can be constructed from the logit life-cycle functions.[18]

The second equation estimated was

$$\text{(4)}\quad \mathrm{PCO}_{it} = \sum_{i=1}^{14} \beta_0^i + \beta_1 \mathrm{UDRS}_{it-1} + \beta_2 \mathrm{SRC}_{it-1} + \beta_3 \mathrm{SRF}_{it-1} + \beta_4 \mathrm{PCOF}_{it} + \sum_{i=1}^{14} \Omega^i \mathrm{RINT}_{t-1}.$$

The results from this equation are contained in Table 4. Again, β_1 is positive and significant at better than the .01 level; while, β_2 is

---

[17] If we include these industries in the sample and do not estimate a life-cycle coefficient for each industry, we are implicitly assuming that PCO is independent of the number of years in franchising.

[18] Consider the convenience store industry in 1984, the representative firm had been in franchising for 12.85 years at that time. Working backward and forward in time from 1984 using the base of $\mathrm{YF}_{84}=12.85$ an entire series for PCOF_{it} can be constructed with the logit life-cycle function.

TABLE 4—SECOND EQUATION,[a] Proportion of Company-Owned Outlets

		$F = 929.12$	$\bar{R}^2 = .9944$			
Industry	Intercept Terms β_0^i	Interest Rates Ω^i	Profitability β_1	Company-Owned Risk β_2	Franchised Risk β_3	Life Cycle β_4
			.00251	−.04800	.02657	−.04886
			(3.093)	(2.526)	(0.692)	(1.775)
Automotive Products and Services	.11998	.00257				
	(12.621)	(2.124)				
Accounting, Credit, Collection Agencies, and General Business Services	.03480	−.00298				
	(4.375)	(1.765)				
Employment Services	.28658	.01365				
	(16.823)	(3.130)				
Printing and Copying Services	.07031	−.00885				
	(5.500)	(3.092)				
Misc. Business Services	.05860	.00114				
	(4.266)	(0.758)				
Construction, Home Improvement, Maint., and Cleaning Services	.06395	−.00504				
	(6.765)	(3.111)				
Convenience Stores	.66334	−.00922				
	(54.732)	(2.643)				
Educational Products and Services	.20205	−.02050				
	(10.376)	(4.785)				
Restaurants	.30648	.00799				
	(28.173)	(6.617)				
Hotels, Motels, and Campgrounds	.17201	−.00259				
	(14.237)	(0.969)				
Recreation, Entertainment, and Travel	4.0834	.00019				
	(4.362)	(0.120)				
Rental Services	.24335	−.00819				
	(25.246)	(3.319)				
Retailing	.30501	.00571				
	(24.150)	(3.940)				
Miscellaneous	.10967	−.00088				
	(8.208)	(0.272)				

[a] All absolute t values are in parentheses.

negative and significant at better than the .02 level. β_3 is positive, but insignificant. The coefficient for PCOF is negative and significant at better than the .08 level, which implies that PCO declines as the firm and the industry matures. The pattern of the results for the real interest rate variable is the same as in equation (1). The joint F test for the interest rate coefficients yields $F = 12.00$. The F value for the entire equation is 929.12.

B. *Risk Effects*

The risk-induced monitoring cost effect suggests a test to discriminate between the two risk effects. If greater variation in sales induces higher monitoring cost, the impact should be asymmetric. Company-owned outlets that achieve sales higher than expected do not require more supervision. More supervision is required when company-owned sales fall below expected sales. Therefore, we divide SRC according to these positive (SRCP) and negative (SRCN) deviations. If the difference between actual sales and expected sales (from the out-of-sample regressions) is positive; then, SRCP = SRC, otherwise SRCP = 0. The same distinction is drawn for SRF; hence, we have SRFP and SRFN. Next, we estimate the following equation

$$(5)\quad PCO_{it} = \sum_{i=1}^{18} \beta_0^i + \beta_1 UDRS_{it-1} + \beta_2^p SRCP_{it-1} + \beta_2^n SCRN_{it-1} + \beta_3^p SRFP_{it-1} + \beta_3^n SRFN_{it-1} + \sum_{i=1}^{18} \Omega^i RINT_{t-1}.$$

Two restrictions were tested: $\beta_2^p = \beta_2^n$ and $\beta_3^p = \beta_3^n$. The test was conducted with and without PCOF and the results were the same for all variables. Therefore, we report only the results from equation (5).

The specific results for relative profitability and the positive/negative risk variables are contained in Table 5. The results are qualitatively the same for al β_0^i and Ω^i, including the joint test of significance for all Ω^i; hence, we do not report those statistics. The F value for the entire equation was 1222.55. The signs and significance levels are consistent with all of the earlier estimates. The F statistic for the hypothesis that $\beta_2^p = \beta_2^n$ is 4.58, which is significant at better than the .04 level. We must reject the null hypothesis and conclude that $\beta_2^p \neq \beta_2^n$. The effects of company-owned risk on short-run PCO are indeed asymmetric. The F statistic for the hypothesis that $\beta_3^p = \beta_3^n$ is 0.81, which is not significant. We cannot reject the null hypothesis in the case of franchised risk. These results are consistent with the existence of risk-induced higher monitoring costs.

If pure risk avoidance does not influence short-run PCO, one would expect the coefficient for SRCP in Table 5 to be insignificant. Since SRCP is defined by the condition that actual company-owned sales are greater than expected in the prior period, one anticipates that the coefficient for SRCP might be positive if risk avoidance does not matter. In fact, the coefficient is negative and significant at better than .01 level. Overall, the evidence is consistent with the conclusion that short-run variations in PCO are influenced by both risk effects.

TABLE 5—THIRD EQUATION

Variable	Coefficient	t Statistic[a]
UDRS	.00301	3.727
SRCP	−.07278	3.877
SRCN	−.04231	2.688
SRFP	.02796	1.094
SRFN	.00805	0.261

[a]Absolute t values.

IV. Implications

Franchising is not a temporary phase on the path to complete ownership integration. It represents a long-run market solution to monitoring and risk diversification problems. As such, attempts to regulate franchising may actually reduce welfare. The franchisor life-cycle study reveals that the proportion of company-owned outlets declines in the long run. The signs and the significance levels of the relative profitability and risk variables in the industry model are consistent with the hypothesis that uncertainty does matter in the franchising choice. Given various sites with different risk and expected return characteristics, the firm will take advantage of the opportunity to diversify the risk and to shed locations with higher risk-induced monitoring costs.

Credit market conditions do appear to influence the short-run mix of company-owned outlets and franchised outlets. In part, the firm's success depends on the exposure provided by expansion. If the firm ceases expansion during tight credit markets, attractive new sites will be lost. If credit is available to the firm and not to the representative franchisee, the firm will expand by company-owned outlets. This captures the site and offers the opportunity to franchise the location at a later date. If credit is available to the franchisee and not to the firm, the firm will expand by franchised outlets. This also captures the site and offers the firm the opportunity to recapture the location as a company-owned outlet when the contract expires. These considerations explain the conventional wisdom among practitioners that capital shortages are the primary motive for franchising.

If franchised outlets require less supervision than company-owned outlets and there are economies of scale in promotion, competitive pressures may force the firm to use franchising more intensively in the short run than it will in the long run. The desire to achieve the minimum efficient scale as quickly as possible suggests that intensive franchising behavior will be observed in the early stages of the firm's life cycle. Economies of scale in monitoring suggest that the firm will use company-owned operations more intensively at the margin once it achieves the minimum efficient scale. The convex structure of many of the life-cycle functions is consistent with these observations.

Finally, the established literature has failed to distinguish between short-run and long-run incentives to franchise. The confusion has led to calls for regulation based on the observed short-run incentives to franchise, when the propriety of such regulation rests on the long-run incentives. Market forces and the long-run incentives to franchise obviate the need for regulation.

APPENDIX I: LIFE-CYCLE FUNCTIONS

TABLE A1—ESTIMATED FUNCTIONS[a]

Industry	YF	YF^2	YF^3	F	R^2
Automotive Products and Services	−.0576	−.0059	.0001	65.769	.7953
	(1.337)	(1.862)	(2.571)		
Accounting, Credit, Collection Agencies, and General Business Services	−.4956	.0303	−.0005	36.979	.8571
	(4.188)	(2.560)	(1.699)		
Employment Services	−.2378	.0164	−.0003	4.436	.2839
	(1.383)	(1.100)	(.992)		
Printing and Copying Services	−.3868	.0132	.0014	33.258	.8432
	(1.833)	(.540)	(1.910)		
Miscellaneous Business Services	−.0885	−.0158	.0005	0.473	−.0556
	(.638)	(.731)	(.594)		
Construction, Home Improvement, Maintenance, and Cleaning Services	−.3319	.0172	−.0003	47.739	.6767
	(4.825)	(2.609)	(2.211)		
Convenience Stores	.0464	−.0003	−.0001	2.718	.1654
	(.350)	(.031)	(.402)		
Educational Products and Services	−.4472	.0412	−.0011	4.062	.3379
	(1.723)	(1.284)	(1.234)		
Restaurants	.0208	−.0038	.0001	60.047	.4090
	(1.117)	(3.470)	(4.264)		
Hotels, Motels, and Campgrounds	−.0463	−.0001	.00002	1.425	.1038
	(.631)	(.028)	(.290)		
Recreation, Entertainment, and Travel	−.1096	−.0079	.0002	7.963	.5987
	(.492)	(.314)	(.393)		
Rental Services	−.5148	.0273	−.0003	6.000	.4839
	(2.621)	(2.441)	(2.395)		
Retailing	−.1715	.0141	−.0002	12.997	.2322
	(3.487)	(4.034)	(4.298)		
Real Estate Services	−.3309	.0508	−.0018	2.833	.2820
	(2.129)	(2.288)	(2.472)		
Consumer Services	−.3653	.0266	−.0004	19.706	.4346
	(5.444)	(4.475)	(4.415)		
Miscellaneous	−.3335	.0202	−.0002	5.564	.5130
	(2.366)	(1.463)	(1.341)		

[a] All absolute t values are in parentheses.

TABLE A2—ASYMPTOTIC 95 PERCENT CONFIDENCE INTERVALS[a] FOR LIFE-CYCLE POINT ESTIMATES

Industry	Lower Bound	Predicted PCO	Upper Bound
Automotive Products and Services	.02	.05	.08
Accounting, Credit Collection, and Business Services	.10	.24	.38
Employment Services	.19	.34	.49
Printing and Copying Services	.01	.02	.03
Miscellaneous Business Services	.19	.40	.61
Const., Home Imp., Maint. and Cleaning Services	.06	.11	.16
Convenience Stores	.32	.43	.55
Educational Products and Services	.00	.21	.45
Restaurants	.33	.36	.39
Hotels, Motels, and Campgrounds	.00	.28	.57
Recreation, Ent., and Travel	.00	.04	.12
Rental Services	.24	.38	.52
Retailing	.51	.55	.59
Real Estate	.15	.38	.61
Consumer Services	.20	.29	.38
Miscellaneous	.00	.70	1.00

[a] Let PCO* be the upper/lower bound. Then, $\text{PCO}^* = \overline{\text{PCO}}^+ \pm 1.96(G\hat{\Sigma}G')^{1/2}$, where $\overline{\text{PCO}}^+$ is the predicted value, G is the gradient vector of $\text{PCO} = F(\beta)$ with respect to the parameters and $\hat{\Sigma}$ is the covariance matrix of the estimated parameters.

APPENDIX II: DATA DEFINITIONS AND SOURCES

Firm Data. The data for franchising organizations were obtained from two sources: *the 1985 Franchise Annual* (FA) and *The Source Book of Franchise Opportunities* (SB). For the 949 franchising organizations contained in the sample I obtained the number of company-owned outlets, the number of franchised outlets, the year in which the firm first began franchising, and the industry to which the firm belongs. The 949 firms represent all of the organizations in either FA or SB that reported these data. Both of these publications contain franchise contract data such as royalties, franchise fees, total investment required, and contract period.

Industry Data. The industry data were assembled from *Franchising in the Economy 1985–1987* (FE). For each of the eighteen industries and for each year from 1969 through 1986, I collected the total number of company-owned establishments in the industry (TCO_{it}), the total number of franchisee-owned establishments in the industry (TFO_{it}), the total industry sales by company-owned establishments (TSCO_{it}), and the total industry sales by franchisee-owned establishments (TSFO_{it}). Hence, $\text{PCO}_{it} \equiv \text{TCO}_{it}/(\text{TCO}_{it} + \text{TFO}_{it})$, $\text{CSO}_{it} \equiv \text{TSCO}_{it}/\text{TCO}_{it}$, and $\text{FSO}_{it} \equiv \text{TSFO}_{it}/\text{TFO}_{it}$. PCO_{it}, CSO_{it}, and FSO_{it} are the statistics for the representative franchisor in the ith industry in the tth year. The ratio of company-owned sales per establishment to franchisee-owned sales per establishment is $\text{RS}_{it} \equiv \text{COS}_{it}/\text{FOS}_{it}$ and the expected sales ratio is $\text{ERS}_{io} \equiv [\sum_{t=-1}^{-5} \text{RS}_{it}]/5$. The company-owned sales risk variable (SRC_{it}) and the franchisee-owned sales risk variable (SRF_{it}) in the industry model are obtained from the five-period sales regression against time as described in the text.

The real interest rate data were obtained from the *Statistical Abstract of the United States 1988*. RINT_t is equal to the federal funds rate in t minus the annual percent change in the GNP deflator for t.

The industry definitions in FA, SB, and FE do not conform exactly. There are many more subclassifications of industries in FA and SB than there are in FE. In most cases the conversions from the classifications in FA and SB into the more general classifications in FE are obvious. For example, there is little ambiguity surrounding Automotive Products and Services, Convenience Stores, Restaurants and Hotels, Motels, and Campgrounds. Judgment was inevitably required in some instances. In any event, industry definitions, copies of all of the data, and the computer programs used in the study are available from the author on request.

REFERENCES

Blair, Roger D. and Kaserman, David L., "Optimal Franchising," *Southern Economic Journal*, October 1982, *49*, 494–505.

______, *Law and Economics of Vertical Integration and Control*, New York: Academic Press, 1983.

Bond, Robert E., *The Source Book of Franchise Opportunities*, Homewood, Ill: Dow Jones-Irwin, 1985.

Brickley, James A. and Dark, Frederick H., "The Choice of Organizational Form: The Case of Franchising," *Journal of Financial Economics*, July 1987, *18*, 401–20.

Caves, Richard E. and Murphy, William F., "Franchising: Firms, Markets, and Intangible Assets," *Southern Economic Journal*, April 1976, *42*, 572–86.

De Alessi, Louis, "Why Corporations Insure," *Economic Inquiry*, July 1987, *25*, 429–38.

Dixon, Edward L., Jr., ed., *The 1985 Franchise Annual*, Lewiston, NY: Information Press, 1985.

Judge, George G., Griffiths, William E., Hill, R. Carter and Lee, Tsoung-Chao, *The Theory and Practice of Econometrics*, New York: Wiley & Sons, 1980.

Klein, Benjamin, "Transaction Cost Determinants of 'Unfair' Contractual Arrangements," *American Economic Review*, May 1980, *70*, 356–62.

______ **and Saft, Lester F.**, "The Law and Economics of Franchise Tying contracts," *Journal of Law and Economics*, May 1985, *28*, 345–61.

Mathewson, G. Frank and Winter, Ralph A., "The Economics of Franchise Contracts," *The Journal of Law and Economics*, October 1985, *28*, 503–26.

Mayers, David and Smith, Clifford W., Jr., "On the Corporate Demand for Insurance," *Journal of Business*, April 1982, *55*, 281–96.

McGuire, E. Patrick, *Franchised Distributors*, New York: The Conference Board, 1971.

Ozanne, Urban B. and Hunt, Shelby D., *The Economic Effects of Franchising*, U.S. Senate, Select Committee on Small Business, Committee Print, 92 Cong., 1st Sess., Washington: USGPO, 1971.

Rubin, Paul H., "The Theory of the Firm and the Structure of the Franchise Contract," *Journal of Law and Economics*, April 1978, *21*, 223–33.

Scott, Frank A., "Vertical Control Via Franchising," Working Paper, February 1987, Department of Economics, University of Kentucky.

U.S. Department of Commerce, Finance and Management Industries Division, *Franchising in the Economy 1985–1987*, Washington: USGPO Annual.

______, Bureau of the Census, *Statistical Abstract of the United States 1988*, 108th ed., Washington: USGPO, 1987.

Irreversible Investment, Capacity Choice, and the Value of the Firm

By ROBERT S. PINDYCK*

Most investment expenditures are irreversible. The firm invests until the full value of an incremental unit of capacity equals its full cost. The former includes the value of the option to not utilize the unit; the latter includes the opportunity cost of sinking resources in the unit. We examine implications for capacity choice, utilization, firm value, and long-run marginal cost. With irreversibility, capacity is smaller, and firm value can be largely attributable to growth possibilities.

Most major investment expenditures are at least partly irreversible: the firm cannot disinvest, so the expenditures are sunk costs. Irreversibility usually arises because capital is industry- or firm-specific, that is, it cannot be used in a different industry or by a different firm. A steel plant, for example, is industry-specific. It can only be used to produce steel, so if the demand for steel falls, the market value of the plant will fall. Although the plant could be sold to another steel company, there is likely to be little gain from doing so, so the investment in the plant must be viewed as a sunk cost. As another example, most investments in marketing and advertising are firm-specific, and so are likewise sunk costs.[1]

The irreversibility of investment has been neglected since the work of Kenneth Arrow (1968), despite its implications for spending decisions, capacity choice, and the value of the firm. When investment is irreversible and future demand or cost conditions are uncertain, an investment expenditure involves the exercising, or "killing," of an option—the option to productively invest at any time in the future. One gives up the possibility of waiting for new information that might affect the desirability or timing of the expenditure; one cannot disinvest should market conditions change adversely. This lost option value must be included as part of the cost of the investment. As a result, the Net Present Value (NPV) rule "Invest when the value of a unit of capital is at least as large as the purchase and installation cost of the unit" is not valid. Instead the value of the unit must *exceed* the purchase and installation cost, by an amount equal to the value of keeping the firm's option to invest these resources elsewhere alive—an opportunity cost of investing.

This aspect of investment has been explored in an emerging literature, and most notably by Robert McDonald and Daniel Siegel (1986). They show that with even moderate levels of uncertainty, the value of this opportunity cost can be large, and investment rules that ignore it will be grossly in error. Their calculations, and those in related papers by Michael Brennan and Eduardo Schwartz (1985) and Saman Majd and Robert Pindyck (1987), show that in many cases projects should be undertaken only when their present value is at least double their direct cost.[2]

---

*Massachusetts Institute of Technology, Cambridge, MA 02139. This is a revised version of my 1986 NBER Working Paper, and is based on research done while I was Visiting Professor of Economics and Fellow, Institute of Advanced Studies, at Tel-Aviv University. I am grateful to the National Science Foundation for research support under grant no. SES-8318990, and to Giuseppe Bertola, Olivier Blanchard, Hua He, Saman Majd, Robert McDonald, Julio Rotemberg, Lawrence Summers, and two anonymous referees for helpful discussions and comments. Any errors are mine.

[1] Partial irreversibility also results from the "lemons'" problem. Office equipment, cars, trucks, and computers are not industry-specific, but have resale value well below their purchase cost, even if new.

[2] Other examples are Ben Bernanke, 1983; Alex Cukierman, 1980; Carliss Baldwin, 1982; and Jeffrey Mackie-Mason, 1988. In the papers by Bernanke and

The existing literature has been concerned with investment decisions involving discrete projects, for example, whether to build a factory. This paper examines capacity choice and expansion, for example, how large a factory to build, and when to expand it. In particular, I focus on the *marginal* investment decision. This provides a simple and intuitively appealing solution to the optimal capacity problem. It also yields insight into the sources of the firm's value, and clarifies the measurement of long-run marginal cost.[3]

A firm's capacity choice is optimal when the present value of the expected cash flow from a marginal unit of capacity just equals the total cost of that unit. This total cost includes the purchase and installation cost, plus the opportunity cost of exercising the option to buy the unit. An analysis of capacity choice therefore involves two steps. First, the value of an extra unit of capacity must be determined. Second, the value of the option to invest in this unit must be determined (it will depend in part on the value of the unit itself), together with the decision rule for exercising the option. In essence, this decision rule is the solution to the optimal capacity problem.

To determine the value of a marginal unit of capacity, we must account for the fact that if demand falls, the firm can choose not to utilize the unit. In effect, a unit of capacity gives the firm an infinite number of options to produce, one for every future time t, each with exercise price equal to production cost, and can be valued accordingly. As we will show, these "operating options" are worth more the more volatile is demand, just as a call option on a stock is worth more, the more volatile is the price of the stock. This suggests that the firm should hold more capacity when future demand is uncertain, but the opposite is true. The reason is that uncertainty also increases the value of the firm's investment options, and hence the opportunity cost of irreversibly investing. Although the value of a unit of capacity increases, this opportunity cost increases even more, so the net effect is to reduce the firm's optimal capacity.[4]

Note that a firm's market value has two components: the value of installed capacity (i.e., the value of the firm's options to utilize some or all of this capacity over time), and the value of the firm's options to add more capacity later. As we will see, numerical simulations suggest that for many firms, "growth options" should account for a substantial fraction of market value, and the more volatile is demand, the larger is this fraction.

Options to productively invest are important assets, which firms hold even if they are price-takers in product and input markets. How do they arise? In some cases it is the result of a patent on a production technology, or ownership of land or natural resources. More generally, a firm's managerial resources, reputation, market position, and possibly scale, all of which may have been built up over time, enable it to productively undertake investments that individuals or other firms cannot undertake.[5]

Cukierman, uncertainty over future market conditions is reduced as time passes, so firms have an incentive to delay investing when markets are volatile (for example, during recessions). In the other papers cited above and in the model I present here, future market conditions are *always* uncertain. Access to the investment opportunity is then analogous to holding a call option on a dividend-paying stock; an expenditure should be made only when the value of the resulting project exceeds its cost by a positive amount, and increased uncertainty raises the incentive to delay the expenditure. Thus the results are similar to those in Bernanke and Cukierman, but for different reasons. Option value also appears in the natural resource context: if future values of wilderness areas and parking lots are uncertain, it may be better to wait before irreversibly paving a wilderness area. See, for example, Claude Henry, 1974.

[3] The analysis in this paper is closely related to that in Giuseppe Bertola, 1987, who independently developed a model of capacity choice similar to mine, using a different solution approach.

[4] In Andrew Abel, 1983, and Richard Hartman, 1972, uncertainty over future prices leads to an increase in the firm's optimal capital stock when the production function is linear homogeneous. The reason is that the marginal revenue product of capital is a convex function of price, so that as in my model, a marginal unit of capital is worth more when price is stochastic. However, in Abel and Hartman investment is reversible, so the opportunity cost of investing is zero.

[5] The importance of growth options and their implications for the firm's financial structure are discussed in

Section I clarifies the nature of the firm's options to invest and produce, and how they affect its choice of capacity and its market value. Section II then solves the capacity choice problem in the context of a specific model. As the model is developed, a numerical example is used to show how the value of a marginal unit of capital, the opportunity cost of investing, and the firm's optimal capacity depend on current demand and uncertainty over future demand. Sections III, IV, and V use the model to study the value of the firm, the behavior of capacity and capacity utilization over time, and implications for the measurement of marginal cost. Section VI concludes with caveats and limitations.

I. Optimal Incremental Investment Decisions

Consider a firm facing a demand curve that shifts over time stochastically, so that future demands are uncertain. Let θ denote the demand shift parameter, with $\partial P(Q,\theta)/\partial\theta > 0$. Suppose the firm can install units of capital one at a time, at a sunk cost k per unit, whenever it wishes. Letting K be the amount of capital currently in place, we can write the value of the firm, W, as the sum of two parts:

$$W = V(K;\theta) + F(K;\theta). \tag{1}$$

$V(K;\theta)$ is the value of the firm's capital in place, that is, the present value of the expected flow of profits that this capital will generate, given the current value of θ. $F(K;\theta)$ is the value of the firm's "growth options," that is, given that the firm has capital K in place and given the current value of θ, $F(K;\theta)$ is the present value of any additional profits that might result should the firm add more capital in the future, less the present value of the cost of that capital. Note that $F(K;\theta)$ is greater than the present value of the expected flow of net profits from *anticipated* future investments, because the firm is not committed to any investment path.

Equation (1) is just an accounting identity, but we can use it to gain insight into the firm's investment problem by noting that units of capital are installed sequentially. Assuming for now that the units are discrete, we can number them in the order they are installed. Suppose units 1 through n have been installed so far. Then, suppressing θ, we can rewrite (1) by summing the value of each installed unit and the values of the options to install further units:

$$\begin{aligned} W = {} & \Delta V(0) + \Delta V(1) \\ & + \Delta V(2) \cdots + \Delta V(n-1) \\ & + \Delta F(n) + \Delta F(n+1) + \cdots . \end{aligned} \tag{2}$$

Here $\Delta V(j)$ is the value of the $j+1$st unit of capital, that is, *the present value of the expected flow of incremental profits generated by unit* $j+1$. Of course the firm need not always utilize this (or any other) unit of capital. It has the option to utilize it at every point during its lifetime, and $\Delta V(j)$ is equal to the value of this option. Section II shows how $\Delta V(j)$ can be calculated.

The firm must decide whether to install additional capital. Given that n units are in place, $\Delta F(n)$ is the value of the option to buy one more unit, that is, unit $n+1$, at any time in the future. If the firm exercises this option, it pays k and receives an asset worth $\Delta V(n)$. The firm also gives up $\Delta F(n)$, because once exercised, the option is dead—whether or not the firm later buys more capital, it has now paid for unit $n+1$, and cannot disinvest. Hence $\Delta F(n)$ is also a cost of investing in this unit. The full cost of investing is thus $k + \Delta F(n)$, and this must be compared to the benefit $\Delta V(n)$.

Once the firm buys unit $n+1$, it then faces another problem: at what point should it exercise its option, worth $\Delta F(n+1)$, to buy unit $n+2$, which is worth $\Delta V(n+1)$? And so on. These options must be exercised sequentially, so the total value of the firm's

---

Stewart Myers, 1977. A complete model of industry evolution would also describe the competitive processes through which firms obtain these options. Such a model is beyond the scope of this paper.

options to grow is

$$F(n)=\sum_{j=n}^{\infty}\Delta F(j).$$

Letting these incremental units become infinitesimally small, we can write equation (2) for the value of a firm with capital stock K as

$$W=\int_0^K \Delta V(\nu;\theta)\,d\nu+\int_K^{\infty}\Delta F(\nu;\theta)\,d\nu. \tag{3}$$

The firm's optimal capital stock K^* is that which maximizes its net value, $W-kK^*$. Using (3), this maximization gives the following optimality condition that must hold if the firm is investing:

$$\Delta V(K^*;\theta)=k+\Delta F(K^*;\theta). \tag{4}$$

Thus the firm should invest until the value of a marginal unit of capital, $\Delta V(K;\theta)$, is equal to its total cost: the purchase and installation cost, k, plus the opportunity cost $\Delta F(K;\theta)$ of irreversibly exercising the option to invest in the unit, rather than waiting and keeping the option alive.[6]

The firm's investment problem can therefore be solved by calculating $\Delta V(K;\theta)$ and $\Delta F(K;\theta)$, and using (4) to determine the optimal capacity $K^*(\theta)$. In the next section we carry out these steps for a specific model.

II. A Model of Capacity Choice

Consider a firm that faces the following demand function:

$$P=\theta(t)-\gamma Q. \tag{5}$$

(The firm might be a price-taker, in which case $\gamma=0$.) Here $\theta(t)$ evolves according to the stochastic process

$$d\theta=\alpha\theta\,dt+\sigma\theta\,dz, \tag{6}$$

where dz is the increment of a Weiner process, that is, $dz=\varepsilon(t)(dt)^{1/2}$, with $\epsilon(t)$ a serially uncorrelated and normally distributed random variable. Equation (6) says that the current value of θ (and thus the current demand function) is known to the firm, but future values of θ are unknown, and are lognormally distributed with a variance that grows with the time horizon. Thus even though information arrives over time (the firm observes θ changing), future demand is always uncertain.[7]

The firm's cost and production constraints are as follows: *i*) each unit of capital can be bought at a fixed price k per unit; *ii*) each unit of capital provides the capacity to produce one unit of output per time period, so that $Q\le K$; and *iii*) the firm has an operating cost $C(Q)=c_1Q+(1/2)c_2Q^2$. In general c_1 and/or c_2 can be zero, but if $\gamma=0$ (so the firm is a price-taker), we require $c_2>0$ to bound the firm's size.

I assume that the firm starts with no capacity, so at $t=0$ it must decide how much initial capacity to install. Later it might add more capacity, depending on how demand evolves. For simplicity I assume that new capacity can be installed instantly, and capi-

---

[6] Note that $\Delta V(K)$ is *not* the marginal value of capital, as the term is used in marginal q and related models of investment, such as the one in Abel and Olivier Blanchard, 1986. The marginal value of capital is the present value of the expected flow of profits throughout the future from whatever unit of capital is the marginal one, that is,

$$E\int_0^{\infty}[\partial\pi_t(K_t)/\partial K_t]e^{-\mu t}\,dt,$$

where μ is the discount rate. The marginal value of capital thus depends on the firm's capital stock, K_t, or its distribution, at every future t, and so its calculation can be quite difficult. Note how this differs from $\Delta V(K)$, the present value of the expected flow of incremental profits attributable to the $K+1$st unit of capital, which is independent of how much capital the firm has in the future.

[7] Analytic solutions can be obtained for any demand function linear in θ or a power function of θ; I use (5) for simplicity. Also, it is straightforward to also allow for uncertainty over future operating costs. The qualitative results would be the same.

tal in place does not depreciate.[8] Finally, investment is completely irreversible—the firm cannot disinvest. This means there is an opportunity cost to investing. Adding a unit of capacity today precludes the possibility of waiting and instead adding the unit later or not at all.

I make one more assumption: stochastic changes in demand are spanned by existing assets, that is, there is an asset or dynamic portfolio of assets whose price is perfectly correlated with θ. (This is equivalent to saying that markets are sufficiently complete that the firm's decision to invest or produce does not affect the opportunity set available to investors.) This assumption holds for most commodities, which are usually traded on both spot and futures markets, and for manufactured goods to the extent that prices are correlated with the values of shares or portfolios. However, in some cases this assumption will not hold, for example, a new product unrelated to any existing ones.

With the spanning assumption, we can determine the investment rule that maximizes the firm's market value, and the investment problem reduces to one of contingent claim valuation.[9] This provides insight, and avoids assumptions about risk preferences or discount rates. However, as shown in the Appendix, the problem can also be expressed in terms of dynamic programming.[10] If spanning does not hold, dynamic programming can still be used to maximize the present value of the firm's expected flow of profits, using an arbitrary discount rate. (Note that in such cases there is no theory for determining the discount rate; the CAPM, for example, would not hold.) $\Delta V(K)$ and $\Delta F(K)$ would satisfy differential equations nearly the same as those that will be derived below (but containing the discount rate), so the qualitative results will be the same.

Let x be the price of an asset or portfolio of assets perfectly correlated with θ, and denote by $\rho_{\theta m}$ the correlation of θ with the market portfolio. Then x evolves according to

$$dx = \mu x\, dt + \sigma x\, dz,$$

and by the CAPM, its expected return is $\mu = r + \phi\rho_{\theta m}\sigma$, where ϕ is the market price of risk. I will assume that α, the expected percentage rate of change of θ, is *less* than μ. (It will become clear later that if this were not the case, no firm in the industry would ever install any capacity. No matter what the current level of θ, firms would always be better off waiting and simply holding the option to install capacity in the future.) Denote the difference between μ and α by δ, that is, $\delta = \mu - \alpha$.

For purposes of comparison, note that if future demand were certain ($\sigma = 0$), and if $\alpha = 0$, the firm's optimal initial capacity would be $K^*(\theta) = (\theta - c_1 - rk)/(2\gamma + c_2)$, where r is the riskfree rate of interest.[11] Equivalently, the firm should add capacity only if $\theta(K) > (2\gamma + c_2)K + c_1 + rk$. We will see that uncertainty makes the optimal capacity smaller than this.

A. *The Value of a Marginal Unit of Capacity*

To solve the firm's investment problem we first determine $\Delta V(K)$, the present value of the expected flow of profits from an incremental unit of capacity, given a current capacity K. Because the unit need not be utilized, the profit it generates at any future time t is a nonlinear function of θ, which is stochastic:

---

[8] Relaxing these assumptions makes no qualitative difference in the results. In fact, allowing for lead times in the construction and installation of new capacity magnifies the effects of uncertainty, as shown in Majd and Pindyck, 1987.

[9] For an overview of contingent claims analysis and its applications, see John Cox and Mark Rubinstein, 1985; and Scott Mason and Robert Merton, 1985.

[10] The connection between the contingent claims approach that I use and dynamic programming is examined in detail by Bertola, 1987.

[11] If $\alpha > 0$, the optimal initial capacity would be smaller than this. The reason is that the cost of investing is fixed but the payoff is growing, so there is a benefit to delay, and $\Delta F(K) > 0$ for all K.

$$(7)\quad \Delta\pi_t(K)=\max[0,(\theta_t-(2\gamma+c_2)K-c_1)].$$

Thus $\Delta V(K)$ can be written as

$$(8)\quad \Delta V(K) = \int_0^\infty \int_0^\infty \Delta\pi_t(K;\theta)f(\theta,t)\,d\theta e^{-\mu t}\,dt,$$

where $f(\theta,t)$ is the density function for θ at time t, and μ is the risk-adjusted discount rate. It is difficult, however, to evaluate (8) directly. In addition, the discount rate μ might not be known.

Instead we obtain $\Delta V(K)$ by solving the following equivalent problem: What is the value of a factory that produces 1 unit of output per period, with operating cost $(2\gamma+c_2)K+c_1$, which it sells in a perfectly competitive market at a price θ_t, and where the factory can be shut down (temporarily and costlessly) if θ_t falls below the operating cost?[12] The Appendix shows that the solution to this problem, obtained using contingent claim valuation methods or equivalently via dynamic programming, is:

$$(9)\quad \Delta V(K)=\begin{cases} b_1\theta^{\beta_1}; & \theta \le (2\gamma+c_2)K+c_1 \\ b_2\theta^{\beta_2}+\theta/\delta-[(2\gamma+c_2)K+c_1]/r; & \\ & \theta \ge (2\gamma+c_2)K+c_1, \end{cases}$$

where

$$\beta_1=-\frac{(r-\delta-\sigma^2/2)}{\sigma^2}+\frac{1}{\sigma^2}\left[(r-\delta-\sigma^2/2)^2+2r\sigma^2\right]^{1/2}>1$$

$$\beta_2=-\frac{(r-\delta-\sigma^2/2)}{\sigma^2}-\frac{1}{\sigma^2}\left[(r-\delta-\sigma^2/2)^2+2r\sigma^2\right]^{1/2}<0$$

$$b_1=\frac{r-\beta_2(r-\delta)}{r\delta(\beta_1-\beta_2)}\times\left[(2\gamma+c_2)K+c_1\right]^{1-\beta_1}>0$$

$$b_2=\frac{r-\beta_1(r-\delta)}{r\delta(\beta_1-\beta_2)}\times\left[(2\gamma+c_2)K+c_1\right]^{1-\beta_2}>0$$

This solution for $\Delta V(K)$ is interpreted as follows. When $\theta<(2\gamma+c_2)K+c_1$, the unit of capacity is not utilized. Then, $b_1\theta^{\beta 1}$ is the value of the firm's option to utilize the unit in the future, should θ increase. When $\theta \ge (2\gamma+c_2)K+c_1$, the unit is utilized. If, irrespective of changes in θ, the firm had no choice but to continue utilizing the unit throughout the future, the present value of the expected flow of profits would be given by $\theta/\delta-[(2\gamma+c_2)K+c_1]/r$. (Costs are certain and so are discounted at the riskfree rate; future values of θ are discounted at the risk-adjusted rate μ, but θ is expected to grow at rate α, so the effective net discount rate is $\mu-\alpha=\delta$.) However, should θ fall, the firm can reduce output and not utilize this unit of capacity. The value of this option is $b_2\theta^{\beta 2}$.

A numerical example will help to illustrate the model. For this purpose I choose $r=\delta=.05$, $k=10$, $c_1=0$, and either $\gamma=.5$ and

---

[12] The valuation of a factory that can be temporarily shut down has been studied by Brennan and Schwartz, 1985; and McDonald and Siegel, 1985. Observe from equation (7) that the present value of an incremental profit at future time t is the value of a European call option, with expiration date t and exercise price $(2\gamma+c_2)K+c_1$, on a stock whose price is θ, paying a proportional dividend δ. This point was made by McDonald and Siegel, 1985. Thus $\Delta V(K)$, the value of our "equivalent factory," can be found by summing the values of these call options for every future t. However this does not readily yield a closed form solution, and I use an approach similar to that of Brennan and Schwartz, 1985.

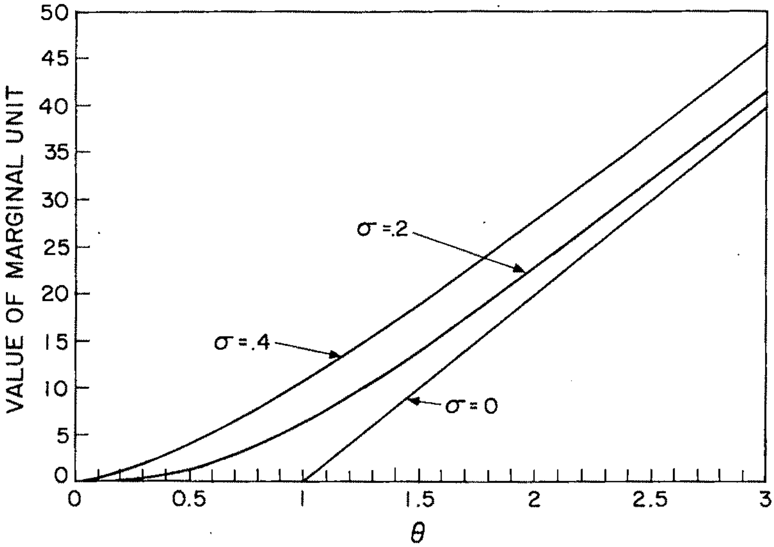


FIGURE 1. THE VALUE OF A MARGINAL UNIT OF CAPACITY, $\Delta V(K)$, AS A FUNCTION OF θ $(K=1)$

$c_2 = 0$, or equivalently $\gamma = 0$ and $c_2 = 1$.[13] I vary θ or K, and consider values of σ in the range of 0 to .4.[14] For comparison, let $\Delta V_0(K)$ denote $\Delta V(K)$ for $\sigma = 0$, so $\Delta V_0(K) = \theta/\delta - [(2\gamma + c_2)K + c_1]/r$ for $\theta \geq (2\gamma + c_2)K + c_1$, and 0 otherwise. In this example, $\Delta V_0(K) = 20(\theta - K)$ for $\theta \geq K$, and 0 otherwise.

Figure 1 shows $\Delta V(K)$ as a function of θ for $K = 1$ and $\sigma = 0$, .2, and .4. Observe that $\Delta V(K)$ looks like the value of a call option —indeed it is the sum of an infinite number of European call options (see fn. 12). As with a call option, $\Delta V(K)$ is increasing with σ, and for $\sigma > 0$, $\Delta V(K) > \Delta V_0(K)$ because the firm need not utilize its capacity. As $\theta \to \infty$, $\Delta V(K) \to \Delta V_0(K)$; for θ very large relative to K, this unit of capacity will almost surely be continuously utilized over a long period of time.

Figure 2 shows $\Delta V(K)$ as a function of K for $\theta = 2$, and $\sigma = 0$, .1, .2, and .4. Because demand evolves stochastically, a marginal unit of capacity has some positive value no matter how large is the existing capital stock; there is always some chance that it will be utilized over any finite period of time. The greater is σ, the more slowly $\Delta V(K)$ declines with K. Also, the smaller is K, the more likely it is that the marginal unit will be utilized, and so the smaller is $\Delta V(K) - \Delta V_0(K)$. When $K = 0$, $\Delta V(0) = \Delta V_0(K)$; with $c_1 = 0$, the *first* marginal unit will always be utilized.

The fact that $\Delta V(K)$ is larger when $\sigma > 0$ might suggest that the firm should hold more

---

[13] With r and δ equal, $\alpha = 0$ if $\sigma = 0$ or if stochastic changes in θ are completely diversifiable (i.e., $\rho_{\theta m} = 0$ so $\mu = r$), but $\alpha > 0$ otherwise. Also, as can be seen from equation (9), $(\gamma = 0,\ c_1 = 1)$ and $(\gamma = .5, c_2 = 0)$ give the same marginal value of capital, and the same optimal behavior of the firm. In the first case the firm is a price-taker but earns inframarginal rent, and in the second case it has monopoly power.

[14] As Zvi Bodie and Victor Rosansky, 1980, show, the standard deviations of annual changes in the prices of such commodities as oil, natural gas, copper, and aluminum are in the range of 20 to 50 percent. For manufactured goods these numbers are usually lower (for example, based on Producer Price Indices for 1948–97, they are 11 percent for cereal and bakery goods, 3 percent for electrical machinery, and 5 percent for photographic equipment). However, variation in the sales of a product for any one company will be much larger than variations in price for the entire industry. Thus a value of σ of 0.2 could be considered "typical" for simulation purposes.

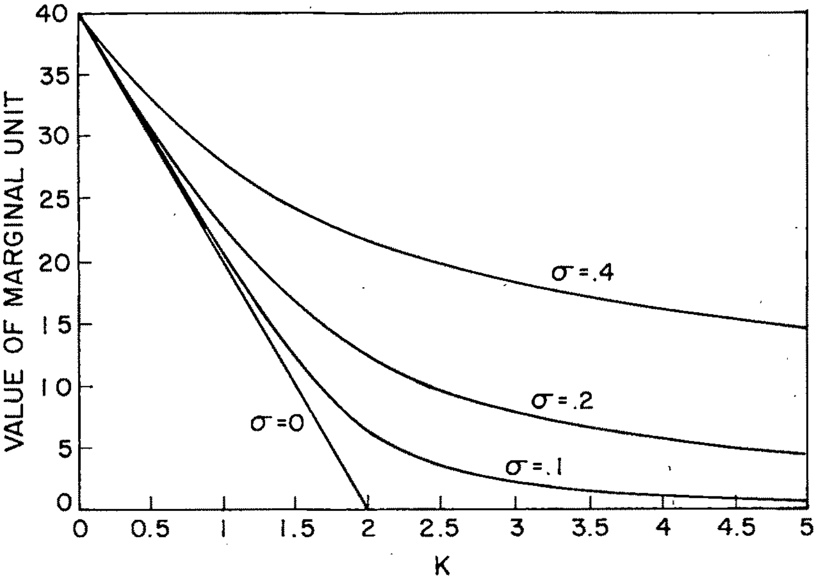


Figure 2. The Value of a Marginal Unit of Capacity, $\Delta V(K)$, as a Function of K ($\theta = 2$)

capacity, but the opposite is true. As shown below, uncertainty also makes the firm's opportunity cost of exercising its option to invest in the marginal unit larger, and by an even greater amount.

B. *The Decision to Invest in the Marginal Unit*

Having valued the marginal unit of capacity, we can now value the firm's option to invest in this unit, and determine the optimal decision rule for investing. In the Appendix it is shown that the value of the firm's option to invest, $\Delta F(K)$, is

$$\Delta F(K) = \begin{cases} a\theta^{\beta_1}; & \theta \leq \theta^*(K) \\ \Delta V(K) - k; & \theta \geq \theta^*(K), \end{cases} \tag{10}$$

where $$a = \frac{\beta_2 b_2}{\beta_1}(\theta^*)^{(\beta_2 - \beta_1)} + \frac{1}{\delta\beta_1}(\theta^*)^{(1-\beta_1)} > 0,$$

β_1, β_2, and b_2 are given under equation (9) above, and $\theta^*(K)$ is the critical value of θ at or above which it is optimal to purchase the marginal unit of capacity, that is, the firm should purchase the unit if $\theta \geq \theta^*(K)$. The critical value $\theta^*(K)$ is in turn the solution to

$$\frac{b_2(\beta_1 - \beta_2)}{\beta_1}(\theta^*)^{\beta_2} + \frac{(\beta_1 - 1)}{\delta\beta_1}\theta^* - \frac{(2\gamma + c_2)K + c_1}{r} - k = 0. \tag{11}$$

Equation (11) can be solved numerically for θ^*, and equation (10) can then be used to calculate $\Delta F(K)$.

Recall our assumption that $\delta > 0$. The reader can verify that as $\delta \to 0$, $\theta^*(K) \to \infty$. Unless $\delta > 0$, the opportunity cost of investing in a unit of capacity always exceeds the benefit, and the firm will never install capacity.[15] Thus if firms in an industry are invest-

---

[15] $\Delta V(K)$ is a function of θ, and if $\delta = 0$, θ is expected to grow at the risk-adjusted market rate. Since the option to invest is perpetual, there would be no gain from installing capacity now rather than later.

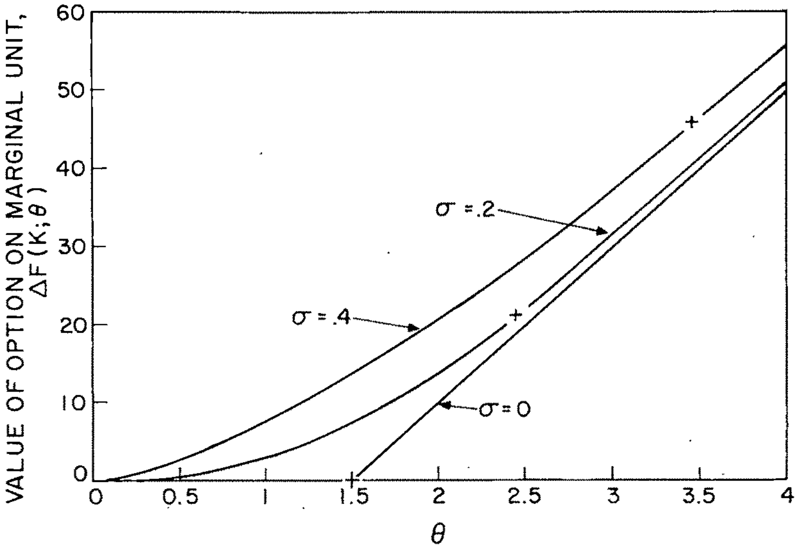


FIGURE 3. THE VALUE OF THE FIRM'S OPTION TO INSTALL A MARGINAL UNIT OF CAPACITY ($K = 1$)

ing optimally and some positive amount of investment is taking place, we should observe $\delta > 0$.

As with a call option on a dividend-paying stock, both $\Delta F(K)$ and the critical value $\theta^*(K)$ increase as σ increases. Figure 3 shows $\Delta F(K)$ as a function of θ for $K = 1$ and $\sigma = 0$, .2, and .4. In each case θ^* is indicated by a "+." When $\sigma = 0$, $\theta^* = 1.5$, that is, the firm should increase capacity only if θ exceeds 1.5. For $\sigma = .2$ and .4, θ^* is 2.45 and 3.44, respectively. The opportunity cost of exercising the firm's option to invest in additional capacity is $\Delta F(K)$, which increases with σ, so a higher σ implies a higher critical value $\theta^*(K)$. Also, it is easily shown that $\theta^*(K)$ is monotonically increasing in K.

C. *The Firm's Optimal Capacity*

The function $\theta^*(K)$ is the firm's optimal investment rule; if θ and K are such that $\theta > \theta^*(K)$, the firm should add capacity, increasing K until θ^* rises to θ. Equivalently we can substitute for $b_2(K)$ and rewrite equation (11) in terms of $K^*(\theta)$, the firm's optimal capacity:

(11′)

$$\frac{r-\beta_1(r-\delta)}{r\delta\beta_1}\theta^{\beta_2}\left[(2\gamma+c_2)K^*+c_1\right]^{1-\beta_2} - \frac{\left[(2\gamma+c_2)K^*+c_1\right]}{r} + \frac{(\beta_1-1)}{\delta\beta_1}\theta - k = 0.$$

Figure 4 shows $K^*(\theta)$ for $\sigma = 0$, .2, and .4. (For many industries .2 is a conservative value for σ—see fn. 14.) Observe that K^* is much smaller when future demand is uncertain. For $\sigma = .4$, θ must be more than three times as large as when $\sigma = 0$ before any capacity is installed.

Another way to see how uncertainty over future demand affects the firm's optimal capacity is by comparing $\Delta F(K)$, the value of the option to invest in a marginal unit, with $\Delta V(K) - k$, the net (of purchase cost) value of the unit. The optimal capacity $K^*(\theta)$ is

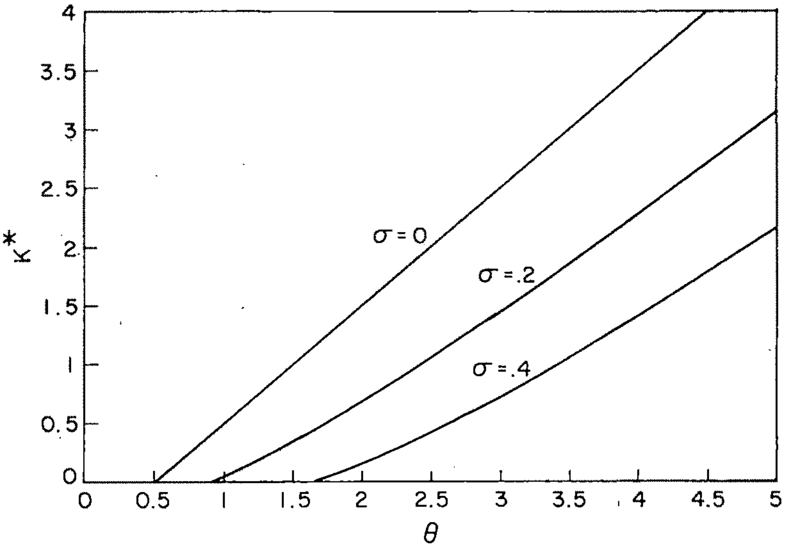


FIGURE 4. OPTIMAL CAPACITY, K^*, AS A FUNCTION OF θ

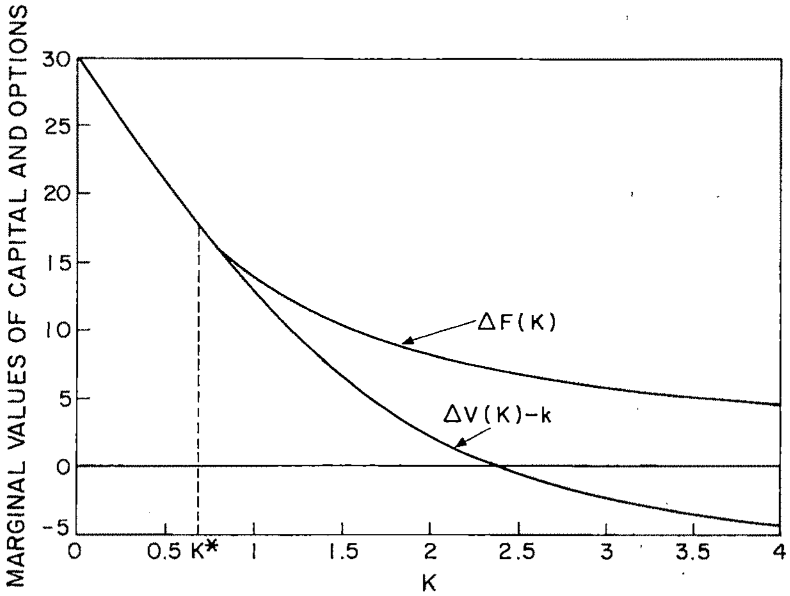


FIGURE 5. THE NET VALUE OF A MARGINAL UNIT OF CAPACITY, AND THE VALUE OF THE OPTION TO INSTALL THE MARGINAL UNIT

the maximum K for which these two quantities are equal. Note from equation (10) that for $\theta \geq \theta^*$, or equivalently, $K \leq K^*$, exercising the option to invest maximizes its value, so that $\Delta F(K) = \Delta V(K) - k$, but for $K > K^*$, $\Delta F(K) > \Delta V(K) - k$, and the option to invest is worth more "alive" than "dead."

This is shown in Figure 5, which plots $\Delta F(K)$ and $\Delta V(K) - k$ as functions of K, for $\theta = 2$ and $\sigma = .2$. Recall that $\Delta V(K)$ is larger when future demand is uncertain. As the figure shows, if the opportunity cost of exercising the option to invest were ignored, that is, if the firm adds capacity until $\Delta V(K) - k = 0$, then capacity would be about 2.3 units (as opposed to 1.5 units when $\sigma = 0$). But at these capacity levels the opportunity cost of investing in a marginal unit exceeds the net value of the unit, so the value of the firm is not maximized. The optimal capacity is only $K^* = .67$, the largest K for which $\Delta F(K) = \Delta V(K) - k$, and the solution to equation (11′).

III. The Value of The Firm

As noted above, $K^*(\theta)$ maximizes the firm's market value, net of cash outlays for the purchase of capital. Recall that the firm's net value as a function of its capacity K is given by:

$$\text{(12)} \qquad \text{Net Value} = \int_0^K \Delta V(\nu)\, d\nu + \int_K^\infty \Delta F(\nu)\, d\nu - kK.$$

Differentiating with respect to K shows that this is maximized when $K = K^*$ such that $\Delta V(K^*) - \Delta F(K^*) - k = 0$.

The value of the firm's installed capacity, $V(K^*)$, is the first integral in equation (12). In Figure 5 it is the area under the curve $\Delta V(K) - k$ from $K = 0$ to K^*, *plus* the purchase cost kK^*. The value of the firm's growth options is the second integral, which in Figure 5 is the area under the curve $\Delta F(K)$ from $K = K^*$ to ∞. As the figure shows, growth options can account for a large portion of the firm's total value.

The sensitivity of firm value and its components to uncertainty over future demand can be seen from Table 1, which shows K^*, $V(K^*)$, $F(K^*)$, and total value for different values of σ and θ. When $\sigma = 0$, the value of the firm is only the value of its installed capacity. Whatever the value of θ, the firm is worth more the more volatile is demand. A larger σ implies a larger value for each unit of installed capacity, and a much larger value for the firm's options to expand. Also, the larger is σ, the larger is the fraction of firm value attributable to its growth options. When $\sigma = .2$ or more, more than half of the firm's value is $F(K^*)$, the value of its growth options. Even when $\sigma = .1$, $F(K^*)$ accounts for more than half of total value when θ is 1 or less. (When demand is currently small, it is the possibility of greater demand in the future that gives the firm much of its value.) And there is always a range of θ for which K^* is zero, so that *all* of the firm's value is due to its growth options.

As mentioned earlier, a σ of .2 or more would not be unusual. Thus an implication of the model is that for many firms, the fraction of market value attributable to the value of capital in place should be one-half or less. A second implication is that this fraction should be smaller the greater is the volatility of market demand. I have not tried to test either of these implications (valuing capital in place is difficult). However, calculations reported by Carl Kester (1984) are consistent with both of them. He estimated the value of capital in place for 15 firms in 5 industries by capitalizing the implied flows of anticipated earnings, and found that it is half or less of market value in the majority of cases. Furthermore, this fraction is only about 1/5 to 1/3 in industries where demand is more volatile (electronics, computers), but more than 1/2 in industries with less volatile demand (tires and rubber, food processing).

IV. The Dynamics of Capacity, Capacity Utilization, and Firm Value

If the firm begins with no capacity, it initially observes θ and installs a starting capacity $K^*(\theta)$. If θ then increases, it will

TABLE 1—VALUE OF FIRM
($c_1 = c_2 = 0, \gamma = .5, r = \delta = .05, k = 10$)

σ	θ	K^*	$V(K^*)$	$F(K^*)$	Value
0	.5	0	0	0	0
	1	0.5	7.5	0	7.5
	2	1.5	37.5	0	37.5
	3	12.5	87.5	0	87.5
	4	3.5	157.5	0	157.5
.1	.5	0	0	0.4	0.4
	1	0.23	4.3	3.4	7.7
	2	1.00	33.1	9.1	42.2
	3	1.82	80.4	20.3	100.7
	4	2.65	147.3	35.5	182.8
.2	.5	0	0	3.1	3.1
	1	0.04	0.8	13.5	14.3
	2	0.67	24.0	49.2	73.2
	3	1.37	67.1	94.6	161.7
	4	2.09	134.1	143.7	277.8
.4	.5	0	0	25.8	25.8
	1	0	0	69.7	69.7
	2	0.15	5.9	182.6	188.7
	3	0.64	36.2	307.2	343.4
	4	1.22	91.3	427.5	518.8

expand capacity accordingly, and the value of the firm will rise. The value of its growth options will also rise, but will become a smaller fraction of total value (see Table 1). However, if θ decreases, the firm will find itself holding more capacity than it would have chosen had the decrease been anticipated. The firm's value will fall, and depending on how much θ decreases, some of its capacity may become idle.

Because capital does not depreciate in this model, the firm's capacity is nondecreasing, but will rise only periodically. The dynamics of capacity are characterized in Figure 6, which shows a sample path for $\theta(t)$, and the corresponding behavior of $K(t)$. (The duration of continuous upward movements in $K(t)$ is exaggerated.) The firm begins at t_0 by installing $K^*(\theta_0) \equiv K_0^*$. Then θ increases until it reaches a (temporary) maximum θ_1 at t_1, and K is increased accordingly to K_1^*. Here is remains fixed until t_2, when θ again reaches θ_1. Afterward K is increased as θ increases, until t_3 when θ begins to decline from a new maximum, and K remains fixed at K_3^*.

Thus an implication of the model is that investment occurs only in spurts, when demand is rising, and only when it is rising above historic levels.[16] Firms usually increase capacity only periodically, and this is often attributed to the "lumpiness" of investment. But lumpiness is clearly not required for this behavior.

Let us now examine the firm's capacity utilization. Clearly during periods of expansion, all capacity will be utilized. When demand falls, however, some capacity may go unutilized, but only if it falls far enough.

If the firm had unlimited capacity it would maximize current profits by setting output at $Q^* = (\theta - c_1)/(2\gamma + c_2)$. However, $K^*(\theta) < (\theta - c_1)/(2\gamma + c_2)$, and as shown in Section II, can be much less even for moderate values of σ. Thus for θ in the range $\underline{\theta}(K) \equiv (2\gamma + c_2)K + c_1 \leq \theta \leq \theta^*(K)$, capacity will remain fixed but will be fully utilized. Capacity will go unutilized only when $\theta < \underline{\theta}(K)$. In Figure 6 this occurs during the intervals (t_a, t_b) and (t_c, t_d).

---

[16] If we allow for depreciation, investment will occur more often and even when demand is below historic highs, but it will still occur in spurts.

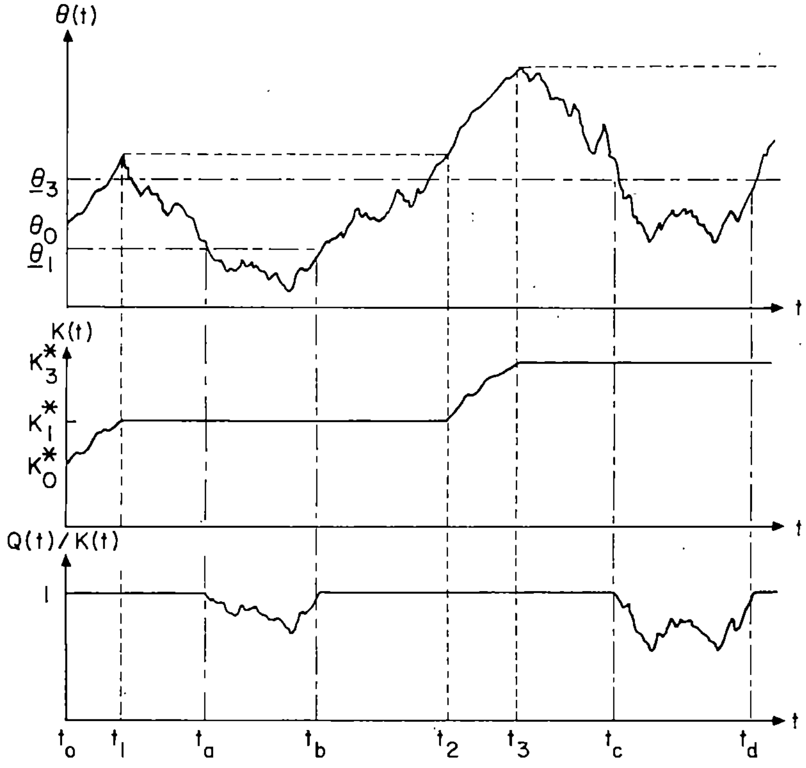


FIGURE 6. REALIZATION OF CAPACITY AND CAPACITY UTILIZATION

The irreversibility of investment induces firms to hold less capacity as a buffer against unanticipated drops in demand. As a result there will be periods of low demands when capacity is fully utilized. A large drop in demand is required for capacity utilization to fall below 100 percent. Most of the time the firm's capacity K will be above $K^*(\theta)$—in Figure 6 exceptions are during the intervals (t_0, t_1) and (t_2, t_3).

The share of the firm's value due to its growth options will also fluctuate with θ. For example, as Table 1 shows, during periods when capacity is growing (so that $K = K^*(\theta)$), this share falls. It also falls when θ is falling and $K > K^*(\theta)$.

V. The Measurement of Long-Run Marginal Cost

The measurement of long-run marginal cost and its relationship to price are important for industry analyses in general, and antitrust applications in particular. When investment is irreversible, traditional measures understate marginal cost and overstate the amount by which it differs from price, even in a competitive market. This problem is particularly severe when product markets are volatile.

Suppose $\sigma > 0$ and $K = K^*(\theta)$. Using equation (9), the optimality condition $\Delta V(K) = k + \Delta F(K)$ can be written as:

$$\text{(13)}\quad \Delta V(K) = b_2\theta^{\beta_2} + \theta/\delta - \left[(2\gamma + c_2)K + c_1\right]/r = k + \Delta F(K)$$

or

$$\text{(13}'\text{)}\quad \theta/\delta - 2\gamma K/r = -\,b_2\theta^{\beta_2} + (c_2K + c_1)/r + k + \Delta F(K).$$

If $\sigma = \alpha = 0$, equation (13′) reduces to a more familiar special case:

$$(14)\quad \theta/\delta - 2\gamma K/r = (c_1 + c_2 K)/r + k.$$

The left-hand side of (14) is capitalized marginal revenue. (Note that θ is capitalized at the rate $\delta = \mu - \alpha$ because it is expected to grow at the rate α but is discounted at the risk-adjusted rate μ; if $\sigma = 0$, $\mu = r$.) The right-hand side is full marginal cost: the capitalized operating cost, plus the purchase cost of a unit of capital. Equation (14) is the usual relation between marginal revenue and marginal cost when the former is increasing at a deterministic rate.

Observe that when $\sigma > 0$, two adjustments must be made to obtain full (capitalized) marginal cost, the RHS of (13′). The first term on the RHS of (13′) is the value of the firm's option to let the marginal unit of capacity go unutilized, and must be subtracted from capitalized operating cost. The last term is the opportunity cost of exercising the option to invest. In our model the last term dominates the first, so that K must be smaller to satisfy (13′), and marginal cost as conventionally measured will understate true marginal cost.

If the firm is a price-taker, $\gamma = 0$ and $P = \theta$. Price will equal marginal cost, if the latter is defined correctly as in (13′). Unfortunately the first and last terms on the RHS of (13′) are difficult to measure, particularly with aggregate data. But if one wishes to compare price with marginal cost, ignoring them can be misleading.[17]

VI. Conclusions

Uncertainty over future market conditions affects investment decisions through the options that firms hold—operating options, which determine the value of capital in place, and options to add more capital, which, when investment is irreversible, determine the opportunity cost of investing. By treating capital as homogeneous and focusing on incremental investment decisions, we have tried to clarify the ways in which uncertainty and irreversibility affect the values of these options, and thereby affect the firm's optimal capacity and its market value.

The assumption that firms can continuously and incrementally add capital, though common in economic models, is extreme. Most investments are lumpy, and sometimes quite so. The opposite extreme assumption is that the firm can build only a single plant, and must decide when to build it (the critical θ^*), and how large it should be (K^*). As sketched out in the Appendix, this problem can also be solved by the methods used in this paper. For our model: *i*) the critical θ^* at which it is optimal to invest increases with σ; *ii*) K^* also increases with σ (operating options are worth more, and there is only one opportunity to invest); and *iii*) for every θ, the value of the firm (the value of the single investment option prior to construction, and the value of the plant after construction) is less than it is when the firm can incrementally invest.

Besides ignoring the lumpiness of investment, the model presented here has other simplifying assumptions. It ignores adjustment costs and delivery lags, and includes only one source of uncertainty. It can be extended to account for these factors, but numerical methods may then be required to obtain solutions. Of course once numerical methods are used, other aspects of the model can also be generalized. For example, demand can be a nonlinear function of θ, or θ could follow some alternative stochastic process, including jumps.

It should be emphasized that the numerical results presented in this paper are based on a specific production technology, and specific functional forms for demand and cost. Also, the assumption that the firm can incrementally invest magnifies the effects of uncertainty, as does the assumption that there is no depreciation. (If capital becomes obsolete rapidly, the opportunity cost of investing will be small.) Thus the quantitative importance of irreversibility and uncertainty may be more limited than the results here would suggest.[18]

---

[17] Robert Hall, 1986, reports that price significantly exceeds marginal cost for most two-digit industries, and finds no explanation for this disparity consistent with competition. Hall's test of marginal cost pricing is based on the relation between the marginal product of labor and the product wage. If firms set marginal operating cost equal to a (constant) proportion of price, his technique will apply, whatever the capital stock. But as shown in Section IV, there can be a wide range of prices for which the firm is capacity constrained, and the ratio of marginal operating cost to price will vary with price.

Subject to these caveats, we find that in markets with volatile and unpredictable demand, firms should hold less capacity than they would if investment were reversible or future demands were known. Also, much of the market value of these firms is due to the possibility (as opposed to the expectation) of increased demands in the future. This value may result from patents and technical knowledge, but it also arises from the managerial expertise, infrastructure, and market position that gives these firms (as opposed to potential entrants) the option to economically expand capacity.

Do firms correctly compute and take into account the opportunity cost of investing when making capacity expansion decisions? Ignoring such costs would lead to overinvestment. John McConnell and Chris Muscarella (1985), have found that for manufacturing firms, market value tends to increase (decrease) when managers announce an increase (decrease) in planned investment expenditures, which is inconsistent with a systematic tendency to overinvest.[19] But there is anecdotal evidence that managers often base investment decisions on present values computed with discount rates that far exceed those that would be implied by the CAPM—diversifiable and nondiversifiable risk are sometimes confused, and an arbitrary "risk factor" is often added to the discount rate. It may be, then, that managers use the wrong method to get close to the right answer.

APPENDIX

Here we derive equation (9) for $\Delta V(K;\theta)$ and equations (10) and (11) for the optimal investment rule and value of the investment option $\Delta F(K;\theta)$.

The value of a marginal unit of capacity, $\Delta V(K;\theta)$, is found by valuing an equivalent "incremental project" that produces 1 unit of output per period at cost $(2\gamma + c_2)K + c_1$, that is sold at price $\theta(t)$, and where the firm can (temporarily and costlessly) shut down if price falls below cost. To value this, create a portfolio that is long the project and short ΔV_θ units of the output, or equivalently the asset or portfolio of assets perfectly correlated with θ. Because the expected rate of growth of θ is only $\alpha = \mu - \delta$, the short position requires a payment of $\delta\theta\Delta V_\theta$ per unit of time (or no rational investor would hold the corresponding long position). The value of this portfolio is $\Phi = \Delta V - \Delta V_\theta \theta$, and its instantaneous return is

$$\text{(A1)} \quad d(\Delta V) - \Delta V_\theta d\theta - \delta\theta\Delta V_\theta\, dt + j[\theta - (2\gamma + c_2)K - c_1]\, dt.$$

The last term in (A1) is the cash flow from the "incremental project"; j is a switching variable: $j=1$ if $\theta(t) \geq (2\gamma + c_2)K + c_1$, and 0 otherwise.

By Ito's Lemma, $d(\Delta V) = \Delta V_\theta d\theta + (1/2)\Delta V_{\theta\theta}(d\theta)^2$. Substitute equation (6) for $d\theta$ and observe that the return (A1) is riskless. Setting that return equal to $r\Phi\, dt = (r\Delta V - r\Delta V_\theta \theta)\, dt$ yields the following equation for ΔV:

$$\text{(A2)} \quad \tfrac{1}{2}\sigma^2\theta^2\Delta V_{\theta\theta} + (r-\delta)\theta\Delta V_\theta + j[\theta - (2\gamma + c_2)K - c_1] - r\Delta V = 0.$$

The solution must satisfy the following boundary conditions:

$$\Delta V(K;0) = 0$$

$$\lim_{\theta \to \infty} \Delta V(K;\theta) = \theta/\delta - [(2\gamma + c_2)K + c_1]/r$$

$$\lim_{\theta \to \infty} \Delta V_\theta(K;\theta) = 1/\delta,$$

and ΔV and ΔV_θ continuous at the switch point $\theta = (2\gamma + c_2)K + c_1$. The reader can verify that (9) is the solution to (A2) and its boundary conditions.

Equation (A2) can also be obtained by dynamic programming. Consider the optimal operating policy

---

[18] Also, the model examines the investment decisions of a single firm, and ignores entry and competition. If rival firms can appropriate the same investment opportunities, δ will be larger, which makes the value of the investment option smaller. Steven Lippman and R. Rumelt, 1985, examine the implications of irreversible investment for a competitive market equilibrium, and Avinash Dixit (1987a,b) studies the implications of sunk costs and stochastic prices for firms' entry and exit decisions, and for the reallocation of capital across sectors.

[19] But they find the opposite true for firms in the oil industry, where there may be a tendency to overinvest in exploration and development.

($j = 0$ or 1) that maximizes the value Φ of the above portfolio. The Bellman equation is

$$\text{(A3)}\qquad r\Phi = \max_{j=(0,1)} \left\{ j[\theta - (2\gamma + c_2) - c_1] - \delta\theta\Delta V_\theta + \frac{1}{dt} E_t d\Phi \right\},$$

that is, the competitive return $r\Phi$ has two components, the cash flow given by the first two terms in the maximand, and the expected rate of capital gain. Expanding $d\Phi = d\Delta V - \Delta V_\theta d\theta$ and substituting into (A3) gives (A2).

Finally, note that ΔV must be the solution to (A2) and the boundary conditions even if the unit of capacity (the "incremental project") did not exist, and could not be included in a hedge portfolio. All that is required is an asset or portfolio of assets (x) that replicates the stochastic dynamics of θ. As Robert Merton (1977) has shown, one can replicate the value function with a portfolio consisting only of the asset x and riskfree bonds, and since the value of this portfolio will have the same dynamics as ΔV, the solution to (A2), ΔV must be the value function to avoid dominance.

Equation (10) for $\Delta F(K;\theta)$ can be derived in the same way. Using the same arguments as above, it is easily shown that ΔF must satisfy the equation

$$\text{(A4)}\qquad \tfrac{1}{2}\sigma^2\theta^2\Delta F_{\theta\theta} + (r-\delta)\theta\Delta F_\theta - r\Delta F = 0$$

with boundary conditions:

$$\Delta F(K;0) = 0$$
$$\Delta F(K;\theta^*) = \Delta V(K;\theta^*) - k$$
$$\Delta F_\theta(K;\theta^*) = \Delta V_\theta(K;\theta^*),$$

where $\theta^* = \theta^*(K)$ is the exercise point, and $\Delta V(K;\theta^*) - k$ is the net gain from exercising. The reader can verify that equations (10) and (11) are the solution to (A4) and the associated boundary conditions.

The assumption that the firm can invest incrementally is extreme. At the opposite extreme, suppose that the firm can build only a single plant, and must decide when to build it and how large it should be. Now the firm has an option, worth $G(K;\theta)$, to build a plant of (arbitrary) size K. Once built, a plant of size K is worth $V(K;\theta) = \int_0^K \Delta V(\nu;\theta)\,d\nu$, where $\Delta V(\nu;\theta)$ is given by equation (9). It is easy to show that if K and θ are chosen optimally to maximize $G(K;\theta)$, then $G(K;\theta)$ must satisfy (A4), with ΔF replaced by G. However, the boundary conditions are now:

$$G(K;0) = 0$$
$$V_K(K^*;\theta^*) - k = 0$$
$$G(K^*;\theta^*) = V(K^*;\theta^*) - kK^*$$
$$G_\theta(K^*;\theta^*) = V_\theta(K^*;\theta^*),$$

where θ^* is again the exercise point, K^* is the optimal plant size, that is, that K which maximizes $[V(K;\theta^*) - kK]$, and $V(K^*;\theta^*) - kK^*$ is the net gain from exercising. Using the first boundary condition, the solution is $G(K;\theta) = a\theta^{\beta_1}$, where β_1 is given under equation (9). The constant a and the critical values θ^* and K^* are determined from the remaining three boundary conditions.

REFERENCES

Abel, Andrew B., "Optimal Investment under Uncertainty," *American Economic Review*, March 1983, *73*, 228–33.

______ **and Blanchard, Olivier J.**, "The Present Value of Profits and Cyclical Movements in Investment," *Econometrica*, March 1986, *54*, 249–73.

Arrow, Kenneth J., "Optimal Capital Policy with Irreversible Investment," in James N. Wolfe, ed., *Value, Capital and Growth, Essays in Honor of Sir John Hicks*, Edinburgh: Edinburgh University Press, 1968.

Baldwin, Carliss Y., "Optimal Sequential Investment When Capital is Not Readily Reversible," *Journal of Finance*, June 1982, *37*, 763–82.

Bernanke, Ben S., "Irreversibility, Uncertainty, and Cyclical Investment," *Quarterly Journal of Economics*, February 1983, *98*, 85–106.

Bertola, Giuseppe, "Irreversible Investment," unpublished manuscript, MIT, November 1987.

Bodie, Zvi and Rosanski, Victor, "Risk and Return in Commodity Futures," *Financial Analysts Journal*, May 1980, *36*, 27–40.

Brennan, Michael J. and Schwartz, Eduardo S., "Evaluating Natural Resource Investments," *Journal of Business*, January 1985, *58*, 135–57.

Cox, John C. and Rubinstein, Mark, *Options Markets*, Englewood Cliffs: Prentice-Hall, 1985.

Cukierman, Alex, "The Effects of Uncertainty on Investment under Risk Neutrality with Endogenous Information," *Journal of Political Economy*, June 1980, *88*, 462–75.

Dixit, Avinash, (1987a) "Entry and Exit Decisions of Firms Under Fluctuating Real Exchange Rates," unpublished manuscript, September 1987.

______, (1987b) "Intersectoral Capital Reallocation Under Price Uncertainty," unpublished manuscript, November 1987.

Hall, Robert E., "The Relation Between Price and Marginal Cost in U.S. Industry," NBER Working Paper No. 1785, January 1986.

Hartman, Richard, "The Effects of Price and Cost Uncertainty on Investment," *Journal of Economic Theory*, October 1972, *5*, 258–66.

Henry, Claude, "Investment Decisions Under Uncertainty: 'The Irreversibility Effect'," *American Economic Review*, December 1974, *64*, 1006–12.

Kester, W. Carl, "Today's Options for Tomorrow's Growth," *Harvard Business Review*, March/April 1984, 153–60.

Lippman, Steven A. and Rumelt, R. P., "Industry-Specific Capital and Uncertainty," unpublished manuscript, September 1985.

Mackie-Mason, Jeffrey K., "Nonlinear Taxation of Risky Assets and Investment, with Application to Mining," NBER Working Paper No. 2631, June 1988.

Majd, Saman and Pindyck, Robert S., "Time to Build, Option Value, and Investment Decisions," *Journal of Financial Economics*, March 1987, *18*, 7–27.

Mason, Scott and Merton, Robert C., "The Role of Contingent Claims Analysis in Corporate Finance," in E. Altman and M. Subrahmanyam, eds., *Recent Advances in Corporate Finance*, Homewood: Richard D. Irwin, 1985.

McConnell, John J. and Muscarella, Chris J., "Corporate Capital Expenditure Decisions and the Market Value of the Firm," *Journal of Financial Economics*, September 1985, *14*, 399–422.

McDonald, Robert and Siegel, Daniel R., "Investment and the Valuation of Firms When There Is an Option to Shut Down," *International Economic Review*, June 1985, *26*, 331–49.

______ **and** ______, "The Value of Waiting to Invest," *Quarterly Journal of Economics*, November 1986, *101*, 707–28.

Merton, Robert C., "On the Pricing of Contingent Claims and the Modigliani-Miller Theorem," *Journal of Financial Economics*, November 1977, *5*, 241–49.

Myers, Stewart C., "Determinants of Corporate Borrowing," *Journal of Financial Economics*, November 1977, *5*, 147–75.

Regulating a Monopolist with Unknown Demand

By TRACY R. LEWIS AND DAVID E. M. SAPPINGTON*

Optimal regulatory policy is derived in a setting where the firm has better knowledge of demand than the regulator. When marginal production costs increase with output, the regulator can induce the firm to use its private information entirely in the social interest. When marginal costs decline with output, however, the regulator is unable to derive any benefit from the firm's superior knowledge, and a single price is established that is invariant to demand.

We examine the design of regulatory policy when the regulated firm has better information about demand than the regulator from the outset of their relationship. Although this seems to be an important and relevant issue to explore, it has been largely overlooked in the literature. Most studies in the recent regulatory literature focus on the setting where a firm's private information concerns its *costs* rather than its demand (see, David Baron and David Besanko, 1984; Baron and Roger Myerson, 1982; Jean-Jacques Laffont and Jean Tirole, 1986; David Sappington, 1983; and Sappington and David Sibley, 1988).

There are a variety of reasons why a regulated firm is likely to have better information about the demand for its product than the regulator.[1] One reason is that the firm generally has superior knowledge of the quality and reliability of the product it is marketing. To the extent that demand is affected by these attributes, the firm is better able to predict the quantity demanded at any price than the regulator. Product quality and reliability are important determinants of demand in industries such as utilities, transportation, and communication, which have traditionally been regulated. Another reason regulated firms will commonly have superior knowledge of demand is because of the significant resources they devote to marketing studies. Although regulatory commissions may have access to the findings of these studies, they generally lack the manpower to fully analyze the findings.[2]

In light of the extensive research that has been devoted to the study of regulatory policy with asymmetric cost information, the corresponding study assuming asymmetric demand information would contribute relatively little if the same qualitative insights arose from the two lines of research. In fact, though, we find striking qualitative differences between the two cases. With asymmetric cost information, significant pricing

---

*University of California at Davis and Bellcore, 435 South Street, Morristown, NJ 07960-1961. We wish to thank two anonymous referees, Robert Russell, William Sharkey, David Sibley, and seminar participants at Boston, Brown, Duke, and Rutgers Universities, The Universities of Iowa and Southern California, and the Federal Trade Commission for helpful comments. The first author gratefully acknowledges financial support from the California Energy Research Group. The views expressed in this paper are not necessarily those of Bellcore.

[1] The fact that regulators commonly believe the firms they regulate to have superior knowledge of demand is reflected in the following excerpt from a report entitled, "Risk, Return, and Ratemaking," by the California Public Utilities Commission's Policy and Planning Division (October 1, 1986): [The Commission must evaluate... arguments with minimal information about a customer's economic situation. Utilities have greater manpower and closer contact with customers, and can probably make better judgments about the conditions customers face.]

[2] Of course, the extent of the information asymmetry in regulated industries is not entirely exogenous. The regulator can devote resources to improving information about demand, and can influence the resources the firm devotes to market studies. Thus, one may view the analysis in this paper as providing some indication of the value of information about product demand in a regulatory setting.

flexibility is generally delegated to the firm. Although the delegation generally results in rents for the firm, social welfare is enhanced when the firm employs its superior technological information to choose prices which track the movement of efficient prices. The regulator also finds it advantageous to induce prices in excess of realized marginal production costs in order to limit the firm's rents. In addition, the regulator will always choose to shut down the firm (i.e., terminate its operations) when realized costs are so high that the maximum net surplus from continued operation is negative. Shutdown may also be induced for particularly high cost realizations even though the potential social gains from production exceed the costs.

With asymmetric demand information, very different conclusions emerge. To illustrate, when marginal costs of production increase with output, the firm commands no rents from its superior demand information, and efficient prices are always implemented.[3] Although pricing authority is delegated to the firm in this setting, the firm can be induced (costlessly) to employ its superior demand information in the social interest. Furthermore, shutdown will occur when and only when production is not in the social interest. On the other hand, if marginal production costs decline with output, the regulator will generally not delegate any pricing authority to the firm.[4] Instead, the regulator will optimally use only his own imperfect knowledge of demand to set a single price that the firm must charge regardless of demand. This procedure results in a price that exceeds the marginal cost of production for small demand realizations and that falls short of marginal cost for large demand. The firm also receives rents under all but some intermediate levels of demand.

With declining marginal costs of production, the possibility of inefficient shutdown also arises. In contrast to the case with asymmetric cost information, if shutdown is induced, it will be implemented for intermediate rather than extreme realizations of the private demand information. Furthermore, the firm may be permitted to produce for some demand realizations even though the maximum net surplus from production is negative for those demand realizations.

The explanation and derivation of these findings proceeds as follows. In Section I, the basic model under consideration is described and a formal statement of the regulator's problem is presented. The first-best outcome is also introduced as a benchmark. The first-best outcome is the one the regulator would implement if he had perfect knowledge of demand. In Section II, it is shown that the regulator can ensure the first-best outcome despite his imperfect demand information when (and only when) marginal production costs rise with output. In Section III, the case of declining marginal costs is examined. It is shown that the regulator will restrict the firm to a single price in this setting. For simplicity, the possibility of shutdown is ignored in Sections I through III.

In Section IV, two extensions of the basic model are considered briefly. The shutdown decision is one of these extensions. The other is to allow for marginal production costs that are not everywhere declining or everywhere increasing with output. The optimal policy when the marginal cost curve is U-shaped is shown to be the natural "combination" of the policies described in Sections II and III.[5] Conclusions are drawn in Section V.

Before proceeding, we briefly distinguish our work from the few studies in the literature that do consider the possibility of asymmetric knowledge of demand. A key distinction is that our analysis is a Bayesian analysis in that the regulator is assumed to have well-defined prior beliefs about the relevant demand parameter. In contrast, a number of analyses (such as Ingo Vogelsang and Jörg Finsinger, 1979; and Mo-Yin Tam, 1985, for example) assume the regulator knows *nothing* about demand (or cost) functions. In this regard, our work is more closely related to the Bayesian analyses of Michael Riordan, 1984, and Daniel Spulber 1988. Spulber considers a bargaining problem in which consumers have private demand information and the firm has private cost information. Riordan's model is more similar to ours in that the firm (alone) has private information only about demand. However, he restricts attention to the case where marginal production costs are everywhere constant in order to focus on the issue of motivating the optimal choice of capacity.

---

[3] Even if marginal costs are strictly increasing in output, average costs may still be falling in the relevant range. Thus, even a natural monopoly firm may have rising marginal costs of production.

[4] This conclusion requires that a common regularity condition be satisfied. The condition will be satisfied if, for example, the firm's private information concerns the position but not the shape of the demand curve.

[5] Another straightforward extension of our findings is to the case where the regulator and firm share the same imperfect knowledge of the cost function. Thus, our conclusions do not hinge on the assumption (maintained for expositional convenience) that costs are known perfectly to both parties.

A third Bayesian model that deserves mention is the general analysis of Roger Guesnerie and Laffont (GL, 1984), which admits the interpretation of private demand information for a regulated firm. Despite a key difference involving a monotonicity condition on the firm's profit function (which (GL) impose but which is not satisfied in our model), many of the proof techniques in (GL) are readily adapted for our purposes. Therefore, we will refer the interested reader at points to (GL) for the proofs of some of our conclusions.

I. Statement of the Model

The regulatory environment under consideration is the following. The cost of producing output q is known to be given by the function $C(q)$. It is also common knowledge that demand for the firm's product is given by $q = Q(p,\theta)$, where p is the unit price for the good. θ captures the firm's private information about consumer demand. Formally, θ is the realization of a random variable described by the density function $g(\theta)$. $g(\theta)$ is assumed to have strictly positive support on the interval $[\underline{\theta},\bar{\theta}]$. $g(\cdot)$ is common knowledge, but the realization of θ is observed only by the firm. The firm observes θ before any interaction with the regulator begins.[6]

In Section III, it will be assumed that the higher the realization of θ, the larger the quantity of output consumers demand at any nonnegative price, p, that is, $Q_\theta(p,\theta) > 0$, where subscripts denote partial derivatives (here and throughout). In this interpretation, θ is a measure of the intensity of demand, or the willingness of consumers to pay for the firm's product. Throughout the analysis, it will be assumed that for all realizations of θ, more output is demanded the lower the regulated price; that is, $Q_p(p,\theta) < 0$. Further, to avoid characterization of local minima rather than maxima, we assume $|C''(Q(p,\theta))| < |Q_p(p,\theta)|^{-1}$ $\forall p,\theta$, when $C''(\cdot) < 0$, where primes denote derivatives. In words, demand curves are assumed to be more steeply sloped than the firm's marginal cost curve when the latter is downward sloping.[7] Finally, for simplicity we abstract from shutdown considerations until Section IV. Thus, we temporarily assume that demand is sufficiently great relative to costs for all realizations of θ that it is preferable to have the firm operate rather than shut down.

The regulator is endowed with the power to establish a unit price for the firm's output and to specify a transfer, T, from consumers to the firm. The funds for the transfer payment might be generated by the fixed component (or "access charge") of a two-part tariff, for example.[8] The regulator can observe (and therefore enforce) the regulated price. The quantity sold at the regulated price is presumed prohibitively costly for the regulator to monitor directly. However, the regulator can still be sure that the firm fulfills its mandate to serve all demand at the regulated price. To do so, the regulator need only invite consumers to report any incident in which they were (either) refused service at the established price (or charged a higher price), and penalize the firm for such rationing.

---

[6] The firm need not know the demand curve it faces exactly for our basic conclusions to hold. The key point is that the firm's knowledge of demand is better than the regulator's knowledge. Exact knowledge of demand does, however, ensure the socially efficient shutdown decision when marginal costs increase with output. (See Section IV.)

[7] This is a standard type of assumption which ensures the regulator's problem is concave. If the relevant demand curve were flatter than the downward-sloping marginal cost curve over some range of output, the possibility arises that in that range, incremental production costs would exceed the incremental benefits to consumers. In particular, production at a level given by the equality between price and marginal cost could provide a minimum of social surplus rather than a maximum.

The assumption that the quantity of output sold by the firm is unobservable is an important one, and so warrants explanation.[9] One simple interpretation of the assumption stems from the observation that there are a variety of attributes of a "unit of output" that are not readily monitored by the regulator. For example, telephone calls of a given duration differ markedly in quality depending upon how quickly the connection is made, the amount of static on the line, etc. Thus, our assumption may be interpreted as a statement that it is prohibitively costly for the regulator to monitor precisely the number of units of "quality-adjusted" output that are sold. Alternatively, even if quality considerations are unimportant, it will generally be very costly for the regulator to collect his own data on sales. And if the regulator relies solely on the firm for information, the firm may find it advantageous to engage in "creative bookkeeping" to misrepresent sales data. In what follows, we abstract from these issues by assuming the regulator cannot observe the firm's sales at all. It is worth noting that when marginal costs rise with output, the inability of the regulator to monitor output is of no consequence. (See Corollary 2 below.)

The regulator's objective is to maximize a weighted average of expected consumers' surplus (S) and profit (Π). $\alpha \in (\frac{1}{2},1]$ will represent the weight on consumers' surplus, and $1-\alpha$ the weight on profit.[10] Formally, the regulator's problem [RP] is the following:

$$\underset{p(\theta),T(\theta)}{\text{Maximize}} \int_{\underline{\theta}}^{\bar{\theta}} \{\alpha[S(Q(p(\theta),\theta))-T(\theta)] + [1-\alpha]\Pi(\theta\,|\,\theta)\}\,g(\theta)\,d\theta$$

subject to:

$$\text{(IR)}\quad \Pi(\theta\,|\,\theta) \geq 0 \;\forall\theta \in [\underline{\theta},\bar{\theta}],$$

$$\text{(IC)}\quad \Pi(\theta\,|\,\theta) \geq \Pi(\hat{\theta}\,|\,\theta) \;\forall\theta,\hat{\theta} \in [\underline{\theta},\bar{\theta}],$$

where

$$\Pi(\hat{\theta}\,|\,\theta) \equiv p(\hat{\theta})Q(p(\hat{\theta}),\theta) - C(Q(p(\hat{\theta}),\theta)) + T(\hat{\theta}).^{11}$$

In the statement of [RP], $p(\theta)$ and $T(\theta)$ are the price and transfer payment that will be implemented when the firm's private demand information is θ. $Q(p(\theta),\theta)$ is the corresponding production level. The optimal regulatory policy can be interpreted as follows. The regulator allows the firm to select a price-transfer pair from a carefully designed menu of such pairs. The incentive compatibility constraints (IC) ensure that the firm prefers the $\{p(\theta),T(\theta)\}$ pair to any other when its private information is θ. The individual rationality constraints (IR) guarantee the firm nonnegative (extranormal) profit from the chosen price-transfer pair. Thus, the firm's voluntary participation in the regulated industry is ensured.

---

[8] Under this interpretation, the fixed component of the two-part tariff (that consumers must pay regardless of the number of units of the product they ultimately purchase) is assumed not to affect demand.

[9] With the cost function known, if output were observable the firm could simply be compensated for its realized costs. Under this policy, the firm would have no incentive to produce a level of output other than the one most preferred by the regulator. Thus, the problem we analyze is only interesting if the firm's sales cannot be observed perfectly by the regulator. For simplicity, we assume they are completely unobservable to the regulator.

[10] Thus, we focus on the case where the regulator "cares more" ($\alpha > \frac{1}{2}$) about consumers than the firm. The regulator's problem is an uninteresting one when $\alpha < \frac{1}{2}$, since unbounded transfers from consumers to the firm are optimal in this case.

[11] Notice that $\pi_\theta(\hat{\theta}\,|\,\theta)$ is positive (negative) when price exceeds (falls short of) marginal cost. This partial derivative will generally change sign in the solution to [RP]. This contrasts with Guesnerie and Laffont (1984, p. 359) where the sign is restricted to be everywhere positive or everywhere negative.

Of central concern is how the optimal regulatory policy is affected by the regulator's imperfect knowledge of demand. As a benchmark, we record the *first-best* policy, which is the one the regulator would implement if the realization of θ were observed publicly.[12]

Definition 1: The first-best policy consists of prices $p^*(\theta)$ and transfer payments $T^*(\theta)$ that satisfy the following properties $\forall\theta\in[\underline{\theta},\bar{\theta}]$:

$$\text{(a)}\quad p^*(\theta)=C'(Q(p^*(\theta),\theta)),\quad\text{and}$$

$$\text{(b)}\quad T^*(\theta)=C(Q(p^*(\theta),\theta)) - p^*(\theta)Q(p^*(\theta),\theta).$$

Thus, the ideal outcome from the regulator's perspective (for any welfare weights $\alpha\in(\frac{1}{2},1]$ and $1-\alpha$) holds the firm to zero profit and implements marginal cost prices. In the next section, we demonstrate that under cost conditions that are commonly assumed in formal analyses, the regulator can implement the first-best policy even if he has no information about demand.

II. Nondecreasing Marginal Costs

In this section, we examine the solution to [RP] when marginal production costs are everywhere nondecreasing. No restrictions are placed on average costs of production. Thus, in particular, average costs of production may be everywhere declining in the relevant range, so the firm may be a natural monopoly. No restrictions are placed on the demand uncertainty either. In particular, $Q_\theta(p,\theta)$ can be both positive and negative in different regions, so demand curves can "cross." The solution to the regulator's problem in this setting is recorded in Proposition 1.

PROPOSITION 1: *If* $C''(q)\geq 0\ \forall q\geq 0$, *then the solution to* [RP] *is the first-best policy.*

PROOF:

It need only be shown that the first-best policy is a feasible solution to [RP] when $C''(q)\geq 0\ \forall q\geq 0$. To do so, consider any two possible distinct realizations of $\theta:\theta_1,\theta_2\in[\underline{\theta},\bar{\theta}]$ with $\theta_2>\theta_1$. Now, for the purposes of this proof only, define $p_i\equiv p^*(\theta_i)$ and $T_i\equiv T^*(\theta_i)$ for $i=1,2$. It remains to show that when $\theta=\theta_i$, the firm will prefer to implement the first-best price and transfer payment $\{p_i,T_i\}$ rather than any other $\{p,T\}$ pair. To do so, it suffices to show that the following two inequalities hold for any $\theta_1,\theta_2\in[\underline{\theta},\bar{\theta}]$:

$$\text{(1)}\quad p_2Q(p_2,\theta_1)-C(Q(p_2,\theta_1))+T_2\leq 0,$$

$$\text{(2)}\quad p_1Q(p_1,\theta_2)-C(Q(p_1,\theta_2))+T_1\leq 0.$$

There are three cases to consider.

Case 1. $Q(p_1,\theta_1)=Q(p_1,\theta_2)$.

In this case, $p_1=p_2$, and $T_1=T_2$; so (1) and (2) are obviously satisfied.

Case 2. $Q(p_1,\theta_1)<Q(p_1,\theta_2)$.

In this case, $p_1\leq p_2$ since $Q_p(p,\theta_2)<0$, with strict inequality if $C''(Q(p_1,\theta_1))>0$. From the definition of first-best prices, whenever $p_2>p_1$, we know $Q(p_2,\theta_2)>Q(p_1,\theta_1)$. Therefore, since $Q_p(p,\theta_1)<0$, we have $Q(p_2,\theta_1)<Q(p_1,\theta_1)<Q(p_2,\theta_2)$.

---

[12] In subsequent discussion, we will refer to the outcome that results under the first-best policy as the first-best outcome. Similarly, we will refer to the prices that are charged in the first-best outcome as first-best prices. Note that in the definition presented here, strictly positive levels of output are assumed to be in the social interest for all demand (θ) realizations. In Section IV, we will consider the possibility that for some sufficiently small realizations of θ, the firm will not produce at all under the first-best policy.

Now, substituting for p_i and T_i, $i=1,2$, from the definition of the first-best policy, inequalities (1) and (2) can be rewritten, respectively, as

$$(3)\quad C'(Q(p_2,\theta_2)) \geq \frac{C(Q(p_2,\theta_2))-C(Q(p_2,\theta_1))}{Q(p_2,\theta_2)-Q(p_2,\theta_1)},$$

$$(4)\quad C'(Q(p_1,\theta_1)) \leq \frac{C(Q(p_1,\theta_2))-C(Q(p_1,\theta_1))}{Q(p_1,\theta_2)-Q(p_1,\theta_1)}.$$

It is apparent that (3) and (4) will hold $\forall\theta_1,\theta_2\in[\underline{\theta},\bar{\theta}]$ with $\theta_2>\theta_1$ if (and only if) $C''(q)\geq 0\ \forall q\geq 0$.

The proof for Case 3, where $Q(p_1,\theta_1)>Q(p_1,\theta_2)$ follows in similar fashion.

To achieve the first-best outcome without knowledge of demand in this case, the regulator need only ask the firm to report the actual demand curve, and then implement the first-best price and transfer payment conditional on the firm's report. The firm will have no incentive to misrepresent its private demand information under this arrangement when marginal costs of production increase with output.[13] Demand exaggeration results in authorization to charge a higher unit price for output; but any revenue enhancement is more than offset by the associated reduction in transfer payment. And should the firm understate demand, the loss in revenue that results from the lower regulated price more than offsets the increment in allowed transfer payment.

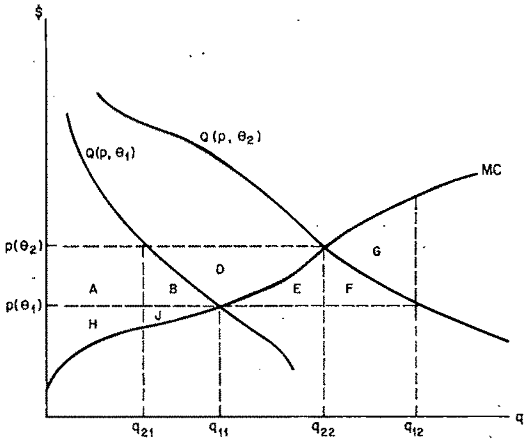


FIGURE 1. INCENTIVES WITH INCREASING MARGINAL COSTS

To examine these effects in more detail, consider Figure 1, which is drawn for the case where $Q_\theta(p,\theta)>0\ \forall p,\theta$. Two representative demand curves are depicted, one for "low" demand (θ_1) and one for "high" demand (θ_2). Quantities are represented on the horizontal axis, with $q_{ij}=Q(p(\theta_i),\theta_j)$, $i,j=1,2$. The first-best policy will implement marginal cost prices, $p^*(\theta_i)$, and award transfer payments, $T^*(\theta_i)(i=1,2)$, which are calculated to provide exactly zero profit to the firm when θ_i is the realized demand parameter. Thus, $T^*(\theta_1)$ will be equal to the fixed costs of production less areas H and J in Figure 1. $T^*(\theta_1)$ will exceed $T^*(\theta_2)$ by the sum of areas A, B, and D.

The conclusion in Proposition 1 can be explained by demonstrating that the firm will strictly prefer to set price $p^*(\theta_i)$ and receive transfer $T^*(\theta_i)$ when $\theta=\theta_i$ rather than select the alternative pair $\{p^*(\theta_j), T^*(\theta_j)\}$, $j\neq i$, $i,j=1,2$. First, consider the change in profit for the firm when $\theta=\theta_1$, but price $p^*(\theta_2)$ is charged. The higher price provides greater revenues on q_{21} units of output. This gain is given by area A in Figure 1. The firm also suffers a direct loss given by area J, since the higher price selected leads to a reduction in demand. In addition, the firm's transfer payment is reduced by the sum of areas A, B, and D when the higher price is selected. Hence, the net loss to the firm from choosing the $\{p^*(\theta_2), T^*(\theta_2)\}$ pair rather than $\{p^*(\theta_1), T^*(\theta_1)\}$ when $\theta = \theta_1$ is the sum of areas B, D, and J.

---

[13] The ensuing discussion is focused on the case where marginal production costs increase strictly with output. When marginal costs are constant, the first-best policy is readily implemented with an allowed price set equal to the constant marginal cost of production and a transfer payment equal to the fixed costs of production. (See Riordan, 1984.)

Now, suppose $\theta = \theta_2$ and the firm chooses to set the lower price $p^*(\theta_1)$ rather than $p^*(\theta_2)$. A direct loss in revenues given by the sum of areas A, B, D, and E results, which more than offsets the increment in transfer payments (areas A, B, and D) that is provided when the lower price is charged. An additional loss given by areas F and G results because the cost of producing the incremental demand $(q_{12} - q_{22})$ exceeds the allowed revenues.

It is also straightforward to verify that the conclusion of Proposition 1 is not altered if the firm cannot be prevented from rationing customers. In Figure 1, when θ_2 is the true demand parameter and the firm chooses to set unit price $p(\theta_1)$, the profit-maximizing output level is q_{11}. But at this level of production, the loss in net revenues exactly offsets the increment in transfer payments that is generated by setting the lower price.

COROLLARY 1: *If* $C''(q) \geq 0\ \forall q \geq 0$, *then the regulator can implement the first-best policy even if he cannot detect rationing of customers.*[14]

Since the first-best outcome is attained in this setting, no strict gains would be realized if it were possible to verify the firm's sales exactly.

COROLLARY 2: *Provided* $C''(q) \geq 0\ \forall q \geq 0$, *any information about realized demand in the industry has no value to the regulator.*

It remains to determine whether the first-best policy might be feasible even when marginal production costs decline with output over some range. As noted in the proof of Proposition 1, this is not the case.

COROLLARY 3: *The first-best policy is a feasible solution to* [RP] *only if* $C''(q) \geq 0\ \forall q \geq 0$.

Hence, it remains to characterize the optimal regulatory policy when marginal production costs decline with output. The analysis begins in Section III.

III. Decreasing Marginal Cost and Single Crossing

The main conclusion to be drawn from Section II is that with rising marginal costs of production, the firm's incentives can be made to coincide with those of society, so the information asymmetry between regulator and firm regarding demand is largely inconsequential. Notice that this conclusion was shown to hold even though virtually no restrictions were placed on the nature of the information asymmetry.

Matters are somewhat more complicated with declining marginal costs. Here, additional structure is required for a complete characterization of the solution to [RP]. To begin, we suppose higher realizations of θ correspond to greater demand at all prices; that is, $Q_\theta(p, \theta) > 0\ \forall p, \theta$. Furthermore, we impose the single-crossing property (SCP), which is commonly assumed in self-selection problems of the type considered here. (See Russell Cooper, 1984, for a survey and general treatment.) Intuitively, the SCP guarantees the firm is systematically more willing (or systematically less willing) to forego transfer payments to obtain a higher unit price the greater is demand. As reported below, the SCP will hold, for example, when increases in θ correspond to parallel outward shifts in the demand curve.

When the SCP holds, it turns out that the firm must be induced to select a higher price the greater is demand under the optimal regulatory policy. (See Lemma 2 below). Thus, an obvious conflict arises between feasible prices and first-best prices (which fall with demand when marginal costs are declining). As is reported in Proposition 2, this conflict is always resolved by instructing the firm to set the same price for its product, independent of its private demand information. In effect, the firm's superior knowledge of demand must be ignored when marginal costs decline with output because it is too costly to make the firm's interests coincide with those of society. Rather than delegate any pricing authority to the firm, the regulator uses his own imperfect demand information to establish the regulated price.

---

[14] The firm will actually be indifferent between implementing the marginal cost price while serving all demand and charging a lower price while limiting output to the profit-maximizing level. Thus, the "loss" that arises when rationing cannot be prevented is that the firm may only weakly prefer to implement the first-best policy, whereas the preference will be strict when rationing can be detected and punished.

To state these conclusions formally, two definitions are helpful.

Definition 2: A price schedule, $p(\theta)$, is *incentive compatible* if it (along with the associated transfer schedule, $T(\theta)$) satisfies the (IR) and (IC) constraints in [RP].

Definition 3: The *single-crossing property* (SCP) holds if the firm's marginal rate of substitution of price for transfer payment $(\Pi_p(p,\theta))$ is monotonic in θ $\forall\theta \in [\underline{\theta},\bar{\theta}]$, and strictly monotonic for some $\theta \in [\underline{\theta},\bar{\theta}]$, where $\Pi(p,\theta) \equiv pQ(p,\theta) - C(Q(p,\theta)) + T(\theta)$.[15]

As noted above, the SCP ensures the firm's preferences vary smoothly and consistently with its private information. Lemma 1 reports that if the firm's private information pertains only to the position of the demand curve and not to its shape, then the SCP is automatically satisfied. (The proof of the Lemma follows from straightforward differentiation of $\Pi(p,\theta)$.)

LEMMA 1: *If* $Q_{p\theta}(p,\theta) = 0$ $\forall\theta \in [\underline{\theta},\bar{\theta}]$ *and* $\forall p \geq 0$, *then* $\Pi_{p\theta}(p,\theta) \geq 0$ $\forall p,\theta$, *so the* SCP *holds.*

Thus, any setting in which higher realizations of θ correspond to parallel outward shifts in demand constitutes a setting where the results cited in this section are valid.[16] Of course, the single-crossing property will also be satisfied more generally.

An implication of the SCP is the following.

LEMMA 2: *If the* SCP *holds, then* $p'(\theta) \geq 0$ $\forall\theta \in [\underline{\theta},\bar{\theta}]$ *in any incentive-compatible price policy.*

The intuition that underlies Lemma 2 is fairly straightforward. (The proof is omitted. See Guesnerie and Laffont, 1984, for the essential details.) To illustrate, take the case where price exceeds marginal production cost. If the firm is to be induced to set a lower price $(p_2 < p_1)$ in this case, it must receive a higher transfer payment $(T_2 > T_1)$. The increase in T required to offset the decrease in revenues will have to be relatively large when demand is high $(\theta = \theta_2)$. However, for smaller realizations of demand $(\theta_1 < \theta_2)$, the same drop in price results in a smaller decrease in revenues. Therefore, if the firm (weakly) prefers $\{p_2, T_2\}$ to $\{p_1, T_1\}$ when $\theta = \theta_2$, the preference will be strict when $\theta = \theta_1$. Consequently, the price charged by the firm in any incentive-compatible policy cannot decline with demand.

But when the regulated price must rise as demand increases to ensure incentive compatibility, price is proceeding in a direction opposite to that of the first-best price. Thus, allowing the firm to use its private knowledge of demand to select the market price becomes very costly...so costly, in fact, that no delegation of pricing authority is allowed.

---

[15] As its name implies, the SCP guarantees that any two iso-profit curves for the firm in $p-T$ space, each corresponding to a different realization of θ, intersect at most once. The additional structure implied by the SCP facilitates the characterization of incentive-compatible price policies: provided the (IC) constraints in [RP] are satisfied locally, they will also be satisfied globally when SCP holds.

[16] One setting in which $Q_{p\theta}(p,\theta) = 0$ $\forall p,\theta$ is the following. Suppose the utility function $H(\cdot)$ of every consumer is separable in income (y) and the quantity (q) of the regulated commodity consumed, and is of the form: $H(q,y,\theta) = h(y) + z(q+\theta)$, where $h(\cdot)$ and $z(\cdot)$ are increasing functions of their arguments. θ might, for example, represent an endowment of the regulated commodity. It is straightforward to verify that the stated property of aggregate demand for the regulated commodity is satisfied in this setting.

PROPOSITION 2: *Suppose* $C''(q) < 0$ $\forall q \geq 0$ *and the* SCP *holds. Then in the solution to* [RP], *the same price* (*and transfer payment*) *will be implemented for all realizations of demand.*

The stark contrast between Propositions 1 and 2 warrants emphasis. (The proof of Proposition 2 is in the Appendix.) When marginal production costs rise with output, the firm is readily motivated to use its private knowledge of demand to implement the outcome that is socially most preferred. On the other hand, when marginal costs decline with output in the setting under consideration, it is so costly to induce the firm to use its superior knowledge in the social interest that any attempt to do so is abandoned: the regulator relies solely on his imperfect demand information to structure regulatory policy.[17]

The single price that is implemented by the regulator depends upon his prior beliefs, upon the firm's cost structure, and upon α. In order to limit the firm's rents, the price generally differs from the price that maximizes expected total surplus. The extent and nature of the difference can be shown to depend upon the third derivative of the firm's cost function.[18]

IV. Extensions of the Model

To facilitate an intuitive understanding of the essential elements of the optimal regulatory policy, a number of simplifying assumptions were maintained in the preceding sections. In particular, marginal costs were assumed to be everywhere declining or everywhere increasing. Also, the possibility of shutting the firm down altogether was ignored. The purpose of this section is to extend the analysis to consider the possibilities of U-shaped marginal cost curves and shutdown.[19]

We begin by examining the optimal regulatory policy when marginal costs of production initially decline over some range of output ($q \in [0, q^L)$) and then rise for higher output levels ($q \in (q^L, \infty)$). We continue to assume the firm's cost structure is common knowledge, and the single-crossing property (SCP) holds. For simplicity, the possibility of shutdown is not considered yet.

The solution to [RP] in this case is illustrated in Figure 2.[20] As is apparent, the optimal price schedule $p^R(\theta)$ is a "combination" of the schedules that are optimal when marginal costs are everywhere declining and when they are everywhere increasing. A constant price, $\bar{p}$, will be charged whenever the realized demand curve intersects the marginal cost curve in its declining range (i.e., for $\theta \in [\underline{\theta}, \theta^L]$, where $p^{*\prime}(\theta) \lesseqgtr 0$ as $\theta \lesseqgtr \theta^L$). This same price will also be implemented when the point of intersection lies in the region where marginal costs begin to increase (i.e., where $p^{*\prime}(\theta) > 0$). However, after some point ($\tilde{\theta}$) in the region where marginal costs (and thus first-best prices) are rising, the first-best prices will be charged.[21]

---

[17] A similar conclusion is reported in Guesnerie and Laffont (1984). The authors present conditions under which a social planner will "ignore" any productivity information possessed by workers in a labor-managed public firm. When these conditions are satisfied, the social planner will simply mandate a fixed level of labor supply for all possible realizations of productivity.

[18] Suppose expected total surplus is a single-peaked function of price; and suppose $Q_{p\theta}(p,\theta) = 0$ $\forall p, \theta$. Then when $C'''(q) > 0$ $\forall q$, the regulator will set price above the level (p^w) that maximizes expected total surplus. The optimal regulated price will be below p^w when $C'''(q) < 0$. And if production costs decline linearly with output, the regulator will implement price p^w. These conclusions hold regardless of the regulator's preferences concerning the distribution of surplus between consumers and the firm (i.e., they hold $\forall \alpha \in (\frac{1}{2}, 1]$).

[19] As noted in fn. 5, the conclusions derived above also carry over in straightforward fashion to the case where the cost function is not known with certainty, but where the regulator and firm share the same imperfect knowledge of costs.

[20] A formal proof of the results that underlie Figure 2 is omitted, as the techniques employed are analogous to those explored by Guesnerie and Laffont (1984).

[21] Note that it is possible to have $\tilde{\theta} = \bar{\theta}$. This is most likely to be the case when $\bar{\theta} - \theta^L$ is small (so that marginal costs decline with output over most of the relevant region) and when $p^*(\bar{\theta}) - p^*(\theta^L)$ is small (so that the ascent of marginal costs is not steep in the region where they are rising). When $\tilde{\theta} = \bar{\theta}$, $p^R(\theta) = \bar{p}$ $\forall \theta \in [\underline{\theta}, \bar{\theta}]$, as in the case where marginal costs are everywhere declining.

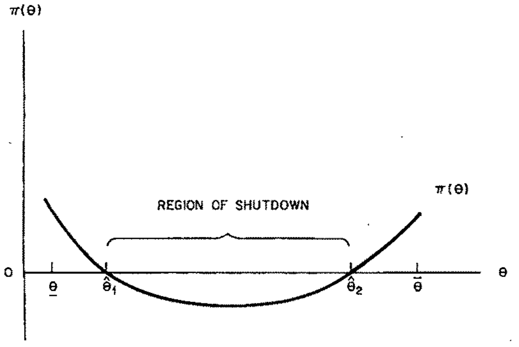


FIGURE 2. U-SHAPED MARGINAL COSTS

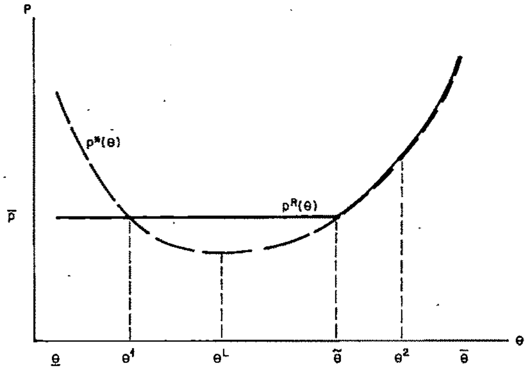


FIGURE 3. DECLINING MARGINAL COSTS AND SHUTDOWN

In the region $[\tilde{\theta}, \bar{\theta}]$ where first-best prices are charged, the firm's profit will be invariant to the realization of θ, as was the case in Section II. However, to induce the firm to set the higher first-best price rather than $\bar{p}$ when $\theta \in [\tilde{\theta}, \bar{\theta}]$, the firm must be afforded the rents it could secure if $\tilde{\theta}$ were the realized demand parameter. Hence, the transfer payment to the firm for setting a price above $\bar{p}$ includes $\Pi(\tilde{\theta} \mid \tilde{\theta})$.

The second extension of the basic model we consider is the shutdown decision. Consider, first, the case in which marginal production costs are known to be everywhere nondecreasing with output. In this case, recall, the firm earns no profit if it operates. Hence, if it is provided with exactly zero profit when it shuts down, the firm will have no incentive to exaggerate demand to avoid being shut down; nor will it gain by terminating operations when production is in the social interest. Thus, when marginal productions costs increase with output, the first-best outcome can be ensured while delegating all pricing and operating decisions to the firm.[22]

Now consider the case where marginal production costs decline with output and the single-crossing property is satisfied. It is not difficult to show that in this case, the regulator will continue to offer a single price and transfer payment for all realizations of demand, even when shutdown may result. Therefore, since the firm's profit increases with θ at a rate directly proportional to the difference between price and marginal cost (see equation (A1) in the Appendix), profit will decline with θ for small demand realizations, and increase with θ for large demand realizations. Consequently, the firm's profit will be minimized, and hence the (IR) constraints will bind, at *intermediate* realizations of θ. Therefore, whenever the solution to [RP] imposes shutdown, production will be absent for some intermediate realizations of demand, as illustrated in Figure 3.[23]

There are a number of interesting implications of this fact. First, the regulator generally cannot induce the firm to terminate operations only when it is socially desirable to do so (i.e., when θ is sufficiently small that the net surplus from operation is negative); additional shutdown may be unavoidable for realizations under which production is socially desirable. Consequently, because this undesired but unavoidable shutdown can be very costly, the fact that net social losses will result from operation for small realizations of θ is not sufficient to ensure shutdown will occur for these realizations. On the other hand, circumstances may arise in which shutdown is induced even though the firm generates positive net social surplus for all demand realizations.[24] Although shutdown does impose some direct losses in this setting, it also reduces the rents the firm can command when it does operate. This latter effect guarantees that if the expected losses from terminating operations are sufficiently small (as when $g(\theta)$ is small for intermediate values of θ), then some degree of shutdown will be optimal. In summary, we have:

---

[22] We make the common assumption here that when it is indifferent among actions, the firm chooses in the social interest.

[23] The fact that shutdown occurs for *intermediate* rather than extreme realizations of θ warrants emphasis, in part because it is in contrast to the findings in related studies in the regulated literature. (See, for example, Baron and Myerson (1982) and Guesnerie and Laffont (1984).) The difference stems from the aforementioned monotonicity of the firm's profit function in θ that is imposed in other studies, but is not imposed here. (Recall fn. 11.)

PROPOSITION 3: *The fact that the net surplus from efficient operation is strictly negative for some realizations of* θ *is necessary and sufficient for shutdown to be implemented for those realizations when* $C''(q) \geq 0$ $\forall q \geq 0$. *It is neither necessary nor sufficient for shutdown when* $C''(q) < 0$ $\forall q \geq 0$. *In this latter case, if shutdown does occur, it occurs for intermediate demand realizations* $[\hat{\theta}_1, \hat{\theta}_2] \subset [\underline{\theta}, \bar{\theta}]$.[25]

V. Conclusions

We have analyzed the optimal regulatory policy when the firm is endowed with superior information about demand, but the regulator and firm share the same technological information. The optimal policy was shown to be sensitive to the firm's technology and to the nature of the demand uncertainty. When marginal costs of production increase with output, the information asymmetry about demand is inconsequential for the regulator. He can (costlessly) induce the firm to employ its superior knowledge in the social interest, thereby implementing the first-best outcome.

On the other hand, when marginal costs decline with output under the demand conditions we have posited, no pricing authority is delegated to the firm.[26] The regulator uses his own imperfect information exclusively to establish the regulated price. In the case of declining marginal costs, the fact that the maximum net surplus from production under the realized demand curve is negative is neither necessary nor sufficient for shutdown of the regulated firm. And when shutdown is optimal, it will generally occur for intermediate rather than extreme realizations of demand.

An important implication of these findings is that the optimal regulatory policy will differ in important qualitative dimensions when the firm is privately informed about demand rather than about costs. Thus, before attempting to draw any general conclusions about appropriate regulatory policy in the presence of asymmetric information, it is imperative that the nature of the information asymmetry be carefully identified.

An important possibility in practice is that the firm will have better information about both its demand and cost functions. It can be shown that if the firm has better information about both demand and the fixed costs of production, but the constant marginal cost of production is common knowledge, the optimal regulatory policy will be as described by Baron and Myerson (1982): price will be set equal to marginal cost, so the only inefficiency that will arise comes in the form of shutting down the firm in cases where total surplus is strictly positive. Thus, the presence of asymmetric information about demand does not alter or seriously complicate optimal regulatory policy in this case because the demand and cost uncertainty are effectively "separable." Absent this separability, however (as when there is asymmetric information both about demand and about marginal production costs), interesting interactions between the two types of uncertainty arise. (For details, see Lewis and Sappington (1988).)

---

[24] Recall that similar predictions are derived from regulatory models with cost uncertainty. See, for example, Baron and Myerson (1982).

[25] It is also possible to show that any region of shutdown $[\hat{\theta}_1, \hat{\theta}_2]$ will be a nondegenerate interval. To see this, note from equation (A1) in the Appendix that if shutdown occurred for a single realization of θ, say $\hat{\theta}^m$, then $p(\hat{\theta}^m) = C'(Q(p(\hat{\theta}^m), \hat{\theta}^m)$, $\Pi(\hat{\theta}^m) = 0$, and $\Pi'(\hat{\theta}^m) = 0$ would all be true. Thus, since shutdown first occurs at a point of efficient operation, social losses from shutdown initially rise at an infinite rate. And since $\Pi'(\theta) = 0$ at $\theta = \hat{\theta}^m$, there is initially no reduction in the firm's expected rents from shutdown. Therefore, since the optimal extent of shutdown is that at which the expected marginal gains and losses are equated, shutdown must occur over a nontrivial interval if it occurs at all.

[26] If the firm's private information concerns both the position and shape of the demand curve in such a way that the SCP is not satisfied, then the optimal regulatory policy may entail prices which decline with demand when marginal production costs decline with output. Consequently, it may be possible to induce the firm to use its private information to set prices which vary with θ in the direction that first-best prices vary with θ. Thus, some delegation of pricing authority may be optimal. A complete characterization of the optimal regulatory policy in this case is problematic because the (IC) constraints in [RP] will generally not hold globally even though they hold locally.

APPENDIX

PROOF OF PROPOSITION 2.

From Lemma 2, $p'(\theta) \geq 0\ \forall \theta \in [\underline{\theta}, \bar{\theta}]$ in any feasible solution to [RP]. Let $p(\theta)$ represent the proposed solution to [RP], and let $p^*(\theta)$ represent the first-best price schedule. Also let $\theta^0 \in [\underline{\theta}, \bar{\theta}]$ represent the value of θ at which the two schedules, $p(\theta)$ and $p^*(\theta)$, intersect. Since $C''(q) < 0\ \forall q \geq 0$, the schedules intersect at most once on the domain $[\underline{\theta}, \bar{\theta}]$. We will return to consider the possibility that they do not intersect at all. The proof is completed for the case where the schedules intersect by showing that the regulatory policy $p^R(\theta) = p^*(\theta^0) = p(\theta^0)$ and $T^R(\theta) = -p(\theta^0)Q(p(\theta^0), \theta^0) + C(Q(p(\theta^0), \theta^0)\ \forall \theta \in [\underline{\theta}, \bar{\theta}]$ is a feasible schedule that is strictly preferred by the regulator to $p(\theta)$.

Since $\{p^R(\theta), T^R(\theta)\}$ consists of a single price and transfer payment, it is obviously incentive compatible. Now, using techniques that are standard in the literature (see, for example, Baron and Myerson (1982), Cooper (1984), and Guesnerie and Laffont (1984)), it is readily shown that in any feasible solution to [RP], $\{p(\theta), T(\theta)\}$, that satisfies the (IC) constraints,

$$\text{(A1)}\quad \Pi'(\theta) = [p(\theta) - C'(Q(p(\theta), \theta))] \times Q_\theta(p(\theta), \theta) \quad \forall \theta \in [\underline{\theta}, \bar{\theta}].$$

Therefore, the proposed regulatory policy also satisfies (IR) because $\Pi(\theta^0) \equiv \Pi(\theta^0 \mid \theta^0) = 0$ and $p^R(\theta) < C'(Q(p^R(\theta), \theta))$ for $\theta \in [\underline{\theta}, \theta^0)$ and $p^R(\theta) > C'(Q(p^R(\theta), \theta))$ for $\theta \in (\theta^0, \bar{\theta}]$.[27]

Next note that by construction, the gap between price and marginal cost is (weakly) reduced under the proposed $p^R(\theta)$ schedule. This has two effects. First, expected total surplus is increased. Second, from (A1), the rate at which the firm's realized profit increases with reductions in θ below θ^0 and with increases in θ above θ^0 is everywhere smaller under $p^R(\theta)$ than under $p(\theta)$. Hence, with expected total surplus increased and the firm's share of the total reduced under the $p^R(\theta)$ schedule, it will provide a larger value for the objective function in [RP], contradicting the presumption that $p(\theta)$ is a solution to [RP].

Finally, notice that if the proposed solution $p(\theta)$ lies everywhere above (below) the first-best price schedule, $p^*(\theta)$, the preceding arguments carry through in analogous fashion with $p^R(\theta) = p^*(\underline{\theta})(p^*(\bar{\theta}))\ \forall \theta \in [\underline{\theta}, \bar{\theta}]$. For an alternative proof of a similar result, see Guesnerie and Laffont (1984).

REFERENCES

Baron, David and Besanko, David, "Regulation, Asymmetric Information, and Auditing," *Rand Journal of Economics,* Winter 1984, *15,* 447–70.

______, **and Myerson, Roger,** "Regulating a Monopolist with Unknown Costs," *Econometrica,* July 1982, *50,* 911–30.

Cooper, Russell, "On Allocative Distortions in Problems of Self-Selection," *Rand Journal of Economics,* Winter 1984, *15,* 568–77.

Guesnerie, Roger and Laffont, Jean-Jacques, "A Complete Solution to a Class of Principal-Agent Problems with an Application to the Control of a Self-Managed Firm," *Journal of Public Economics,* July 1984, *25,* 329–69.

Laffont, Jean-Jacques and Tirole, Jean, "Using Cost Observation to Regulate Firms," *Journal of Political Economy,* June 1986, *94,* 614–41.

Lewis, Tracy and Sappington, David, "Regulating with Unknown Demand and Cost Functions," *Rand Journal of Economics*, Winter 1988, *19*, forthcoming.

Riordan, Michael, "On Delegating Price Authority to a Regulated Firm," *Rand Journal of Economics,* Spring 1984, *15,* 108–15.

Sappington, David, "Optimal Regulation of a Multiproduct Monopoly with Unknown Technological Capabilities," *Bell Journal of Economics,* Autumn 1983, *14,* 453–63.

______, **and Sibley, David,** "Regulating Without Cost Information: The Incremental Surplus Subsidy Scheme," *International*

---

[27] Standard analyses also reveal that if the SCP holds, any regulatory policy that satisfies (A1) also satisfies the (IC) constraints in [RP].

Imperfect Information and Staggered Price Setting

By LAURENCE BALL AND STEPHEN G. CECCHETTI*

Many Keynesian macroeconomic models are based on the assumption that firms change prices at different times. This paper presents an explanation for this "staggered" price setting. We develop a model in which firms have imperfect knowledge of the current state of the economy and gain information by observing the prices set by others. This gives each firm an incentive to set its price shortly after other firms set theirs. Staggering can be the equilibrium outcome. In addition, the information gains can make staggering socially optimal even though it increases aggregate fluctuations.

In many Keynesian models of the business cycle, firms change prices at different times. Even if individual prices change frequently, this "staggered" price setting leads to inertia in the aggregate price level, which causes nominal disturbances to have large and long-lasting real effects (see Olivier Blanchard, 1983, 1986; and John Taylor, 1980). A frequent criticism of this research is that the timing of price changes is treated as exogenous. The models show that staggering has important macroeconomic effects, but they do not explain why staggering occurs. In fact, if the firms in these models are allowed to choose when to change prices, all firms change them simultaneously.[1]

This paper attempts to strengthen the foundations of Keynesian models by presenting an explanation for staggered price setting. We develop a model in which firms have imperfect knowledge of the current state of the economy and gain information by observing the prices set by others. This gives each firm an incentive to set its price shortly after other firms set theirs. Staggering can be the equilibrium outcome.[2]

The argument that imperfect information can lead to staggered price setting is an old one. Arthur Okun (1981), for example, argues that firms' concern for relative wages combined with their ignorance of one another's plans leads to staggering. Okun describes a hypothetical economy in which all firms set wages on January 1 of each year. He then speculates:

> [T]he inability of firms to assess relative wage prospects would destabilize the synchronized situation. Every employer would like to make a decision in full light of decisions that others had made, but would also like to respond promptly. So an employer would want to move a bit behind the schedule followed by the others. As a result, some employer would decide to shift the wage adjustment date to February 1, in order to observe what all the other employers had done. Others would also

---

*Graduate School of Business Administration, New York University; and Ohio State University and NBER, respectively. This is a revised version of NBER Working Paper No. 2201, April 1987. We are grateful for many helpful suggestions from Olivier Blanchard, Larry Jones, David Romer, Julio Rotemberg, Matthew Shapiro, Michael Whinston, the referees, and seminar participants at Columbia, North Carolina State, Ohio State, and the Federal Reserve Bank of Kansas City.

[1] The result that synchronization is the equilibrium timing is apparently a long-standing folk theorem. It is demonstrated formally for the Blanchard model in this paper and in Laurence Ball and David Romer (1987b).

[2] Other explanations for staggering are presented by Gary Fethke and Andrew Policano, 1984,1986; Eric Maskin and Jean Tirole, 1985; Michael Parkin, 1986; and Ball and Romer, 1987b. One prominent informal explanation is that firms change prices at different times simply because they face different shocks and different costs of price adjustment. If this were the entire explanation, however, a large enough nominal shock would cause all prices to adjust immediately. Only moderate shocks would have real effects.

> want to make such a move, but obviously everyone cannot exercise the preference to bat last. The likely result of this "time-location" problem is analogous to that of some spatial location problems. It generates a tendency to spread the distribution of wage-adjustment dates around the calendar.
> [p. 95]

While Okun discusses wages, his point applies equally well to price setting in general. Similar discussions of imperfect information and staggering appear in recent macroeconomics textbooks (for example, Robert Hall and Taylor, 1988, ch. 15).

Despite their popularity, these explanations for staggering have never been formalized. This paper shows that adding imperfect information to Blanchard's (1986) model of monopolistically competitive price setters can create endogenous staggering. In our model, each firm's desired price depends separately on local and aggregate demand shocks, but the firm observes only their sum. As a result, if all firms change prices at the same times, each faces a version of Robert Lucas's (1973) signal extraction problem. But when price decisions are staggered, a price setter observes prices set recently by other firms. These provide information about the previous price setters' estimates of the underlying shocks, and this information improves the current firm's estimates.

We use the model to address two questions. First, when does imperfect information lead to staggering? We characterize the conditions under which staggering and synchronization are stable Nash equilibria. Second, can staggering be socially optimal? When information considerations are absent, staggering is Pareto inferior to synchronization because it leads to price-level inertia, which exacerbates business cycles. But in our model, staggered price setting helps firms set prices closer to full-information levels and may lead to a net improvement in welfare.

Much of our analysis, including our welfare results, depends only on the presence of a firm-specific and an aggregate demand shock that price setters cannot distinguish. It turns out, however, that these two shocks are *not* sufficient to make staggering a stable decentralized equilibrium. Stable staggering arises only when we assume that each firm belongs to a small "neighborhood" of firms—an industry or geographic area—and that there is a third demand shock that is common to neighbors. Throughout, we compare the implications of the model with and without neighborhood shocks.[3]

The remainder of the paper consists of six sections. Section I presents the model and Section II describes the information gains from staggering. Section III derives the behavior of the aggregate price level under staggering and synchronization. As in previous work, staggering leads to price-level inertia. Section IV uses the results of the earlier sections to determine which of the two regimes is socially optimal, and Section V determines when each is a stable decentralized equilibrium. Finally, Section VI discusses generalizations and offers conclusions.

Our results illustrate the complementarity of new classical and new Keynesian macroeconomic models. Lucas's framework of imperfect information provides a foundation for Blanchard's model of staggered price setting. At the same time, the possibility of staggering makes more plausible the idea that information imperfections are an important source of aggregate fluctuations. In actual economies, these imperfections appear short-lived. For example, data on the U.S. price level are available with approximately a one-month lag. Short information lags can lead to staggering, however, and staggering causes nominal shocks to have long-lasting real effects.

---

[3] The presence of shocks specific to sectors of the economy is also crucial in Fethke and Policano's (1984, 1986) models of staggering. In their work, however, the sectors must be large, and price setting is synchronized within sectors (for example, all manufacturing firms change prices in even periods and all service firms in odd periods). In our model, the economy contains many small sectors, and (as we show below) price setting can be staggered within sectors. We view our model as more realistic. Note also that the source of staggering in Fethke-Policano is very different from the source in this paper. In their models, staggered wage setting arises because adjustment of one sector's wage when other wages are fixed causes beneficial changes in relative wages. In our model, by contrast, the relative price movements caused by staggering are undesirable (again, we believe that this is realistic). Most important, in our model the benefit from staggering is that firms learn about the current values of demand shocks by observing one another's prices. This benefit is not present in Fethke-Policano, because there firms have complete information about current shocks.

I. The Model

Our model combines elements of the models in Blanchard, 1986; Blanchard and Nobuhiro Kiyotaki, 1987; and Blanchard and Stanley Fischer, 1988. The economy contains a large number of firms that sell differentiated products. Each firm fixes its nominal price for two periods. Departing from previous work, we assume that each firm chooses the timing of its price changes; that is, it chooses whether to set its price in even- or odd-numbered periods. In addition, a firm faces several demand shocks and cannot fully distinguish between them.

Part A of this section describes tastes, technology, and price setting. Here, the only departure from previous papers is the addition of product-specific taste shocks to utility; these taste shocks are the source of local demand shocks. Part B describes the structure of shocks and agents' information, which are the crucial parts of the model.

A. *Tastes, Technology, and Price Setting*

The economy consists of N agents, where N is a large number. Each agent is a producer-consumer: he uses his own labor to produce a differentiated good, then sells this product and purchases the products of all other agents. Agents take one another's prices as given. With these assumptions, we suppress the labor market and focus on price setting in the goods market. (In future sections, we generally refer to agents as "firms.")[4]

Omitting time subscripts, agent i's utility function is

$$(1)\quad U_i = C_i - \frac{\varepsilon\gamma}{\varepsilon-1}L_i^{\gamma},$$

$$C_i = N\left[\frac{1}{N}\sum_{j=1}^{N}\left(\phi_j C_{ij}\right)^{(\varepsilon-1)/\varepsilon}\right]^{\varepsilon/(\varepsilon-1)},$$

where C_{ij} is agent i's consumption of agent j's product; C_i, is an index of agent i's total consumption; L_i is agent i's labor supply; ϕ_j is a shock to tastes for agent j's product; ε is the elasticity of substitution between any two goods ($\varepsilon > 1$); and γ measures the degree of increasing marginal disutility of labor ($\gamma > 1$). The coefficient multiplying L_i is chosen for convenience.

Agent i has a linear production function

$$(2)\qquad Y_i = L_i,$$

where Y_i is agent i's output. A money demand function determines the relation between real balances and total spending on goods

$$(3)\qquad Y = M/P,$$

where

$$Y = \frac{1}{N}\sum_{j=1}^{N}\frac{P_j Y_j}{P};\quad \text{and}$$

$$P = \left[\frac{1}{N}\sum_{j=1}^{N}\left(P_j/\phi_j\right)^{1-\varepsilon}\right]^{1/(1-\varepsilon)};$$

M is the money stock; P_j is the price of agent j's product; and P is the price index for consumption (agent i must spend P to obtain one unit of C_i).[5]

---

[4] Our model of producer-consumers is a simplification of Blanchard and Kiyotaki, 1987, which contains both goods and labor markets. Ball and Romer, 1987a, 1987b, study a similar economy of producer-consumers.

[5] Our qualitative results are robust to natural modifications of (1)–(3). For example, we can introduce decreasing returns in production (this has the same effects as increasing the taste parameter γ); we can add real money balances to the utility function (as in Blanchard and Kiyotaki) rather than simply assuming (3); or we can introduce risk aversion in consumption (as in Ball and Romer, 1987a).

The utility function determines the demand for agent i's product as a fraction of aggregate spending. Writing this relation in logs,

$$y_i - y = -\varepsilon(p_i - p) + u_i, \tag{4}$$

where $y = \ln(Y)$, $p = \ln(P)$, and $u = (\varepsilon - 1)\ln(\phi)$.[6] Combining (3) and (4) yields product demand

$$y_i = (m - p) - \varepsilon(p_i - p) + u_i, \tag{5}$$

where $m = \ln(M)$. According to (5), demand for agent i's product depends on three variables: real money, which determines aggregate demand; the agent's relative price (with an elasticity determined by the substitutability of goods); and a shock determined by tastes for the agent's product. (Below we assume that the shock u_i has both firm-specific and "neighborhood" components.)

If agent i set his price every period with full information, he would choose the price that maximizes utility, (1), subject to (2), (5), and the constraint that his consumption spending equal his revenue ($PC_i = P_iY_i$). In logs, the utility-maximizing price is

$$p_i^* = v(m + u_i) + (1 - v)p, \tag{6}$$

$$v = \frac{\gamma - 1}{1 + \gamma\varepsilon - \varepsilon}.$$

We assume, however, that agent i fixes his price for two periods. If he sets a price at t, it is in effect at t and $t+1$. This assumption captures the idea that costs of price adjustment lead agents to change prices less frequently than shocks arrive. In principle, we could make the adjustment cost explicit and assume that agents optimize over the frequency of price changes as well as the timing (as in Ball, 1987). But adding another choice variable would complicate the analysis tremendously, and so we treat the frequency of adjustment as exogenous.[7]

Substituting the production function and product demand into (1) yields agent i's utility in terms of his price and the other variables that affect demand (p, m, and u_i). For simplicity, in the analysis below we approximate this indirect utility function by $-(p_i - p_i^*)^2$; that is, we assume that the agent minimizes the squared deviations of his price from the utility-maximizing level. (Because demand is loglinear, $(p_i - p_i^*)^2$ is proportional to $(y_i - y_i^*)^2$, the squared deviation of the agent's output from the utility-maximizing level.)[8]

In choosing a price for t and $t+1$, agent i uses all information available at the end of $t-1$. Thus, ignoring discounting, the agent minimizes the loss function

$$\mathscr{L}_i = E_{t-1}^i\{(x_{it} - p_{it}^*)^2 + (x_{it} - p_{it+1}^*)^2\}, \tag{7}$$

where x_{it} is the log of agent i's price for t and $t+1$, and E_{t-1}^i is the expectation conditional on his information at the end of $t-1$. Minimization of (7) implies the simple price-setting rule

$$x_{it} = \tfrac{1}{2}\{E_{t-1}^i p_{it}^* + E_{t-1}^i p_{it+1}^*\}. \tag{8}$$

Finally, to simplify aggregation we approximate the log of the price index in (3) by[9]

---

[6] For details of the derivation of (4), see the analogous derivations in Blanchard and Fischer and in Blanchard and Kiyotaki.

[7] In the concluding section, we discuss the implications of assuming that prices are fixed for more than two periods. Note that the cost of price adjustment need not be large for firms to fix prices for two or more periods, because a "period" can be short. As we emphasize throughout the paper, staggering causes nominal shocks to have large real effects even if individual prices change frequently.

[8] Similar quadratic objective functions appear in Jo Anna Gray, 1976; Parkin, 1986; Ball, 1987; and many other papers. Ball and Romer, 1987b, Appendix, study a model of staggered price setting without making this paper's approximation to utility.

$$p_t \simeq \frac{1}{N}\sum_{j=1}^{N} p_{jt}. \tag{9}$$

B. *The Structure of Shocks and Agents' Information*

For realism, we now refer to the model's producer-consumers as "firms."

Both the money stock and local demand are stochastic. m_t follows a random walk

$$m_t = m_{t-1} + \delta_t, \qquad \delta_t \sim N(0, \sigma_m^2). \tag{10}$$

The local demand shock u_i contains both a firm-specific component and a component that is common to firms within a "neighborhood." Specifically, assume that each firm belongs to a neighborhood of n firms, where n is a *small* number. The economy contains a large number of neighborhoods. If firm i is a member of neighborhood I, then

$$u_{it} = \theta_{It} + \theta_{It-1} + \eta_{it}, \tag{11}$$

where θ_I is common to firms within a neighborhood but uncorrelated across neighborhoods, and η_i is uncorrelated across all firms. The neighborhood shock θ and the firm-specific shock η are both white noise with mean zero and variances σ_θ^2 and σ_η^2. Equation (11) implies that the neighborhood shock affects u_i for two periods.[10]

A neighborhood can be interpreted as an industry or geographic area. The neighborhood shock captures the realistic idea that a firm's demand is more highly correlated with the demand of "neighbors" than with the demand of other firms. This will mean that a firm learns more by observing neighbors' prices than by observing prices in the rest of the economy. The presence of neighborhood shocks and the persistence of their effects prove essential to the result, in Section V, that staggered price setting can be a stable decentralized equilibrium. But much of our analysis holds for the special case in which u_i contains only a firm-specific component ($\sigma_\theta^2 = 0$). Since this case is relatively simple, we emphasize it throughout.

Our central assumption is that agents have imperfect information. Firms observe one another's prices when they are set, but the demand shocks θ and η and the money stock (which should be interpreted as nominal aggregate demand) are observed with a lag. Specifically, assume that all shocks are revealed, for example through announcements by the government or trade associations, with a two-period lag. Firms set x_{it} based on information revealed through $t-1$, so they have full information about shocks two periods earlier, at $t-3$. Assuming that shocks are revealed with a two-period lag, rather than a longer one, simplifies the analysis but is not crucial. As will be clear below, a shorter lag would eliminate the information gains from staggering, which are the focus of the paper.[11]

II. The Information Gains from Staggering

A. *Overview*

The remainder of the paper studies the economy described in Section I. We focus on two price-setting regimes: synchronization, in which all firms set prices in even periods; and uniform staggering, in which half set prices in even periods and half in odd periods. When neighborhood shocks are present ($\sigma_\theta^2 > 0$) and price setting is staggered, we assume that each neighborhood is split equally between even and odd periods.[12] This section describes the price-setting problem facing a firm in each regime to show that there are information gains from staggering. Later sections build on this analysis to determine when synchronization and staggering are socially optimal, and when each is a stable decentralized equilibrium.

---

[9] This is a first-order approximation used by Blanchard and Fischer (1988) and many others. The taste shocks in (3) drop out of the first-order approximation because, as we specify below, the mean of the log taste shocks is zero. Again, see Ball and Romer (1987b, Appendix) for the implications of foregoing the approximation.

[10] Our qualitative results are robust to other specifications of the stochastic processes of the shocks. The only essential assumptions are that the aggregate, neighborhood, and firm-specific shocks follow different processes and that the first two are not white noise.

[11] In the United States, the money stock is announced with a very short lag. But if we interpret m_t as nominal aggregate demand (money times velocity), it is realistic to assume a significant lag.

To understand the benefits from staggering, consider the information available to a firm when it sets x_{it}. As we specify above, the firm observes the underlying demand shocks (m, θ, and η) only through $t-3$, but observes all prices, and hence the aggregate price level, through $t-1$. The demand equation, (5), shows that the firm can infer the sum of its demand shocks, $(m+u_i)$, from its own price and sales and the aggregate price level. Since the firm always knows its own price and sales, $(m+u_i)$ is observed through $t-1$.

If price setting is synchronized, $(m+u_i)$ is firm i's only information about shocks since $t-3$. But if price setting is staggered, prices set at $t-1$ reveal further information. As we show below, prices set at $t-1$ depend on $(m_{t-2}+u_{it-2})$, the last composite shock before $t-1$. The local shock u_i averages to zero over the many $t-1$ price setters, so the average of $t-1$ prices reveals m_{t-2}.[13] Thus a firm setting x_{it} under staggering knows the money stock through $t-2$ even though m is directly observed only through $t-3$. (In constrast, under synchronization firms observe prices in effect at $t-1$, but these were all set at $t-2$ based on information about $t-3$. Since $t-3$ shocks are directly observed, the prices reveal nothing new.)

Staggering yields this information gain regardless of whether the local shock is purely firm-specific ($\sigma_\theta^2=0$) or includes a neighborhood component. With neighborhood shocks, staggering has an additional benefit: by observing prices set by neighbors at $t-1$, a firm gains information about its $t-2$ neighborhood shock, θ_{It-2}. This proves useful both in distinguishing m from u_i and in distinguishing the two components of u_i. The complicated details are described below.

The rest of this section formally describes a firm's price-setting problem and its information set. We begin with the simple model without neighborhood shocks and then consider the general model. In each case, we compare the synchronized and uniformly staggered regimes.

B. *The Simple Model*

The Synchronized Regime. If all firms change prices at $t, t+2$, and so on, the aggregate price level does not change in alternate periods. This means that $p_t=p_{t+1}$ and $E_{t-1}^i p_t=E_{t-1}^i p_{t+1}$. Since the money stock follows a random walk, $E_{t-1}^i m_{t+1}=E_{t-1}^i m_t=E_{t-1}^i m_{t-1}$. Finally, since $u_i=\eta_i$ in the simple model and η_i is serially uncorrelated, $E_{t-1}^i u_{it}=0$. Combining these results with firm i's price-setting rule, (8), and the expression for the firm's desired price, (6), leads to

$$x_{it}^S=vE_{t-1}^i m_{t-1}+(1-v)E_{t-1}^i p_t, \tag{12}$$

where x_{it}^S is firm i's price in the synchronized regime.

Equation (12) shows that x_{it}^S depends on firm i's estimates of m_{t-1} and p_t. Since m_{t-3} is announced, estimation of m_{t-1} reduces to estimation of the last two innovations in the money stock, $(\delta_{t-1}+\delta_{t-2})$. As noted above, firm i's sales at $t-1$ and $t-2$ reveal $(m_{t-2}+u_{it-2})$ and $(m_{t-1}+u_{it-1})$. In the simple model, these signals equal $(m_{t-2}+\eta_{it-2})$ and $(m_{t-1}+\eta_{it-1})$. Along with m_{t-3}, they reveal two pieces of information useful in estimating $(\delta_{t-1}+\delta_{t-2})$:

$$z_1=\delta_{t-1}+\delta_{t-2}+\eta_{it-1} \tag{13}$$

and

$$z_2=\delta_{t-2}+\eta_{it-2}. \tag{14}$$

---

[12] One can show that no other pattern of price setting is ever a stable equilibrium.

[13] A firm's local shock is correlated with that of its neighbors because it contains a neighborhood component, θ. But the shock still averages to zero over all $t-1$ price setters, because the economy contains a large number of neighborhoods.

When price setting is synchronized, the firm has no additional information.

Since the δ's are normally distributed, the expectation of their sum conditional on z_1 and z_2 is given by a linear projection

$$(15)\quad E^i_{t-1}(\delta_{t-1}+\delta_{t-2}) = a_1 z_1 + a_2 z_2,$$

where

$$a_1 = \frac{\sigma_m^2 + 2\sigma_\eta^2}{\sigma_m^2 + 3\sigma_\eta^2 + \dfrac{(\sigma_\eta^2)^2}{\sigma_m^2}} \quad \text{and}$$

$$a_2 = \frac{\sigma_\eta^2}{\sigma_m^2 + 3\sigma_\eta^2 + \dfrac{(\sigma_\eta^2)^2}{\sigma_m^2}}.$$

The expectation of m_{t-1} follows immediately:

$$(16)\quad E^i_{t-1}m_{t-1} = m_{t-3} + a_1 z_1 + a_2 z_2.$$

According to (16), expected money depends on the last announced money stock, m_{t-3}, and the noisy information about recent changes in money.

Below we substitute (16) into firm i's price-setting rule to solve for the aggregate price level. But first, we turn to price setting and information in the staggered regime.

The Uniformly Staggered Regime. When half of the firms change prices in each period, p_t does not equal p_{t-1}. Instead of (12), the price-setting rule implied by (6) and (8) is

$$(17)\quad x^U_{it} = vE^i_{t-1}m_{t-1} + \frac{1-v}{2}\left(E^i_{t-1}p_t + E^i_{t-1}p_{t+1}\right);$$

where x^U_{it} is firm i's price in the uniformly staggered regime.

Once again, a firm's price depends on its estimate of m_{t-1}. Crucially, this estimate is better under staggering than under synchronization. As noted above, firms in the staggered regime infer m_{t-2} from prices set at $t-1$. Thus estimation of m_{t-1} reduces to estimation of the single shock δ_{t-1}, not $(\delta_{t-1} + \delta_{t-2})$ as in the synchronized regime.

Formally, we assume that staggered firms observe m_{t-2} and then verify this after solving for the price level. It turns out that each price set at $t-1$ is a linear combination of prices set at $t-2$ and shocks at $t-2$ and $t-3$. Since local shocks average to zero across the economy, the average of prices set at $t-1$ depends only on prices set at $t-2$ and the money stock at $t-2$ and $t-3$. When firm i sets x^U_{it}, it knows prices at $t-2$ and the money stock at $t-3$ (which has been announced). Therefore, the average of prices set at $t-1$ reveals the money stock at $t-2$.

To estimate δ_{t-1}, the unknown part of m_{t-1}, firm i uses

$$(18)\quad z_3 = \delta_{t-1} + \eta_{it-1},$$

which it infers from its $t-1$ sales. Projection of δ_{t-1} onto z_3 leads to

$$(19)\quad E^i_{t-1}m_{t-1} = m_{t-2} + bz_3,$$

where $$b = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_\eta^2}.$$

Not surprisingly, a comparison of (16) and (19) shows that the error in estimating m_{t-1} has smaller variance under staggering than under synchronization. In the simple model, this is the information gain from staggering.

C. *The Neighborhood Model*

When the local shock u_i contains a neighborhood component, the price-setting problem facing firms is more complicated. In addition, there is a second information gain from staggering. This will be important for the result that staggering can be a stable equilibrium. We again consider synchronization and uniform staggering in turn.

The Synchronized Regime. In the neighborhood model, the price-setting rule for synchronized firms implied by (6) and (8) is

$$(20)\quad x_{it}^S = vE_{t-1}^i m_{t-1} + (1-v)E_{t-1}^i p_t^S + \frac{v}{2}E_{t-1}^i\theta_{It-1}.$$

x_{it}^S depends on firm i's estimates of both the aggregate and the neighborhood shock at $t-1$. (It depends on θ_{IT-1} but not the firm-specific shock η_{it-1} because, by (11), only the former affects demand at t.) As in the simple model, the firm observes $(m_{t-1} + u_{it-1})$ and $(m_{t-2} + u_{it-2})$. Along with the announcements of $t-3$ shocks, these reveal

$$(21)\quad z_1^N = \delta_{t-1} + \delta_{t-2} + \theta_{It-1} + \theta_{It-2} + \eta_{it-1}$$

and

$$(22)\quad z_2^N = \delta_{t-2} + \theta_{It-2} + \eta_{it-2}.$$

Firm i estimates θ_{It-1} and the unknown parts of m_{t-1}, δ_{t-1}, and δ_{t-2}, by projecting them on z_1^N and z_2^N:

$$(23)\quad E_{t-1}^i\delta_{t-1} = a_1^N z_1^N + a_2^N z_2^N$$

$$(24)\quad E_{t-1}^i\delta_{t-2} = a_3^N z_1^N + a_4^N z_2^N$$

$$(25)\quad E_{t-1}^i\theta_{It-1} = a_5^N z_1^N + a_6^N z_2^N,$$

where a_1^N through a_6^N depend on the variances of the shocks, σ_m^2, σ_θ^2, and σ_η^2. (The expressions for $a_1^N - a_6^N$ are given in the Appendix.) The firm's estimate of m_{t-1} follows from (23) and (24):

$$(26)\quad E_{t-1}^i m_{t-1} = m_{t-3} + (a_1^N + a_3^N)z_1^N {}_1 + (a_2^N + a_4^N)z_2^N.$$

The Uniformly Staggered Regime. In the neighborhood model, the price-setting rule under staggering is

$$(27)\quad x_{it}^U = vE_{t-1}^i m_{t-1} + \frac{1-v}{2} \times (E_{t-1}^i p_t^U + E_{t-1}^i p_{t+1}^U) + \frac{v}{2}E_{t-1}^i\theta_{It-1}.$$

Again, staggered firms infer m_{t-2} from prices set at $t-1$. Thus they need only estimate δ_{t-1} and the neighborhood shock θ_{It-1}. As in the simple model, one relevant piece of information is

$$(28)\quad z_3^N = \delta_{t-1} + u_{it-1} = \delta_{t-1} + \theta_{It-1} + \theta_{It-2} + \eta_{it-1}.$$

But when neighborhood shocks are present, there is an additional piece of information provided by neighbors' prices. Each price set by a neighbor of firm i at $t-1$, x_{jt-1}, depends on $(m_{t-2} + u_{jt-2})$. Since m_{t-2} is known under staggering, the price reveals $u_{jt-2} = (\theta_{It-2} + \theta_{It-3} + \eta_{jt-2})$. Because $t-3$ shocks are announced, this means that firm i knows $\theta_{It-2} + \eta_{jt-2}$, the sum at $t-2$ of the neighborhood shock and the neighbor's firm-specific shock. Firm i also infers $\theta_{It-2} + \eta_{it-2}$, the $t-2$ sum of the neighborhood shock and its *own* firm-specific shock, from its sales. The firm averages this information over the half of its neighborhood that set prices at $t-1$ and itself—a total of $(n/2)+1$ observations—and obtains

$$(29)\quad z_4^N = \theta_{It-2} + \bar{\eta}_{it-2},$$

where $\bar{\eta}_{it-2}$ is the average of the η_{jt-2}'s and η_{it-2}. The η's do not average to zero, because the neighborhood size, n, is small. (This will be important for some of our results.)

Observation of z_4^N is the second information gain from staggering. (Under synchronization, no prices are set at $t-1$, and so firms cannot construct z_4^N.) Intuitively, z_4^N is useful as an estimate of θ_{It-2}. Because of the persistence of the neighborhood shock, θ_{It-2} is part of u_{it-1}. Thus information about θ_{It-2} helps a firm disentangle δ_{t-1} from u_{it-1}.

Formally, firm i estimates the unknown shocks δ_{t-1} and θ_{It-1} by projecting them on z_3^N and z_4^N:

$$(30)\quad E_{t-1}^i\delta_{t-1} = b_1^N z_3^N + b_2^N z_4^N$$

and

$$(31)\qquad E^i_{t-1}\theta_{It-1} = b^N_3 z^N_3 + b^N_4 z^N_4,$$

where the b^N's depend on the variances of the shocks and the neighborhood size, n (see the Appendix). The neighborhood size matters because $\bar{\eta}$ is the average of η over $(n/2)+1$ observations, so $\mathrm{Var}(\bar{\eta}) = (\sigma^2_\eta/(n/2+1))$. An increase in n reduces the variance of $\bar{\eta}$, making z^N_4 a better estimate of θ_{It-2}.

Firm i's estimate of m_{t-1} follows from $E^i_{t-1}\delta_{t-1}$:

$$(32)\quad E^i_{t-1}m_{t-1} = m_{t-2} + b^N_1 z^N_3 + b^N_2 z^N_4.$$

As in the simple model, information is better under staggering than under synchronization: the errors in estimating m_{t-1} and θ_{It-1} have smaller variances.

III. The Aggregate Price Level

The next step is to solve for the behavior of the aggregate price level under synchronization and staggering. Since neighborhood shocks average to zero across the economy, the results for the simple model and the neighborhood model are similar. We describe only the general case.

A. *The Synchronized Regime*

Let x^S_t be the average across firms of x^S_{it}. When all prices are set at t (even), x^S_t is the aggregate price level at t and $t+1$. Only monetary shocks affect x^S_t, because all other shocks average to zero. We solve for x^S_t by the method of undetermined coefficients. Guess the general solution

$$(33)\qquad p^S_t = x^S_t = m_{t-3} + \pi_1\delta_{t-1} + \pi_2\delta_{t-2}(t \text{ even}).$$

To solve for the π's, substitute (33) into the price-setting rule (20). This yields x^S_{it} in terms of expectations of monetary and neighborhood shocks. Evaluate the expectations using (23)–(25). Aggregating the result yields p^S_t in terms of monetary shocks alone. Finally, set the coefficients on the shocks equal to the π's in (33) and solve for

$$\pi_1 = \frac{v\left(a^N_1 + a^N_3\right) + \frac{v}{2}a^N_5 + (1-v)\,a^N_3\pi_2}{1-(1-v)\,a^N_1}$$

and

$$\pi_2 = \frac{v\left(a^N_2 + a^N_4\right) + \frac{v}{2}a^N_6 + \left[(1-v)\,a^N_2 + 1\right]\pi_1}{1-(1-v)\,a^N_4}.$$

(See the Appendix for details of these calculations.)

B. *The Uniformly Staggered Regime*

In the staggered regime, half the prices in effect at t are set at t and half are set at $t-1$. Thus the aggregate price level is

$$(34)\qquad p^U_t = \tfrac{1}{2}\left(x^U_t + x^U_{t-1}\right),$$

where x^U_t is the average of prices set at t. To solve for x^U_t, guess the general solution

$$(35)\qquad x^U_t = \lambda_0 x^U_{t-1} + \lambda_1 m_{t-1} + (1-\lambda_0-\lambda_1)m_{t-2}.$$

To find the coefficients, substitute (34) and (35) into the price-setting rule (27) and calculate expectations using (30) and (31). Aggregating and setting the coefficients in the resulting expression equal to the λ's yields[14]

$$\lambda_0 = \frac{1-\sqrt{v}}{1+\sqrt{v}} \quad \text{and}$$

$$\lambda_1 = \frac{\left[4v + 2v\left(b^N_3/b^N_1\right) + (1-v)(1-\lambda_0)\right]b^N_1}{4 - b^N_1(1-v)(2+\lambda_0)};$$

$$\lambda_0, \lambda_1 > 0; \quad \lambda_0 + \lambda_1 \leq 1.$$

---

[14] There are two solutions for x^U_t. We choose the stable one, $0 < \lambda_0 < 1$. Note that x^U_{t-1} depends on x^U_{t-2},

(Again, see the Appendix.) Combining (34) and (35) yields the solution for the aggregate price level:

$$(36)\quad p_t^U = \lambda_0 p_{t-1}^U + \tfrac{1}{2}\left[\lambda_1 m_{t-1} + (1-\lambda_0)m_{t-2} + (1-\lambda_0-\lambda_1)m_{t-3}\right].$$

Equation (36) shows that staggering leads to price-level inertia—that is, to slow adjustment of the price level to shocks. The degree of inertia depends on λ_0. Perhaps surprisingly, it is independent of the variances of local and aggregate shocks. Inertia is greatest (λ_0 is largest) when v is small. In turn, v is small when ε is large (product demand is highly elastic) and when γ is small (the marginal disutility of labor increases slowly).

IV. The Optimal Timing of Price Changes

A. *The Approach*

This section computes a firm's loss under synchronization, $\mathcal{L}^S$, and its loss under staggering, $\mathcal{L}^U$, to determine which regime is socially optimal. Synchronization is optimal if $\mathcal{L}^S < \mathcal{L}^U$.[15] To compute each loss, we combine the price-setting rule, (20) for synchronization and (27) for staggering, with the solutions for the price level, (33) and (36), and the formula for a firm's desired price, (6). This yields expressions for $(x_{it} - p_{it}^*)$ and $(x_{it} - p_{it+1}^*)$, the deviations of a firm's price from the desired level at t and $t+1$. Squaring these results and taking expectations yields the values of the loss function, (7), in each regime. The Appendix presents details of these calculations and the final expressions for $\mathcal{L}^S$ and $\mathcal{L}^U$.

B. *Results: The Simple Model*

We first present results for the model without neighborhood shocks. In this case, the relative sizes of $\mathcal{L}^S$ and $\mathcal{L}^U$ depend on two parameters: $\rho = \sigma_u/\sigma_m$, the ratio of the standard deviations of local and aggregate shocks; and v. Numerical calculations determine the locus of (v, ρ) pairs for which $\mathcal{L}^S = \mathcal{L}^U$. This is plotted as the top line in Figure 1. Below the line, $\mathcal{L}^S < \mathcal{L}^U$, so synchronization is optimal. Above the line, staggering is optimal.

To understand this result, recall that staggering has both benefits and costs. The benefit is improved information. There are two costs: price-level inertia, which exacerbates fluctuations in real aggregate demand; and the unintended movements in relative prices that occur when some prices adjust while others are fixed. Since there are both advantages and disadvantages, it is not surprising that staggering is optimal for some parameter values but not for others.

Figure 1 shows that synchronization is optimal when either v or ρ is small. As noted above, v is small when product demand is highly elastic (ε is large). Elastic product demand implies that the fluctuations in relative prices caused by staggering are very costly. In addition, firms reduce these costly fluctuations by setting prices close to those set in the previous period; thus staggering leads to a high degree of inertia and large fluctuations in aggregate demand. These large costs imply that staggering is undesirable despite the information gains.

Turning to the role of ρ, note that if $\rho = 0$, so there are no local shocks, then synchronization is optimal. This case is essentially the Blanchard model. Information is perfect, and so there is no information gain from staggering. On the other hand, as long as ρ is

---

m_{t-2}, and m_{t-3}. Since a firm setting x_{it}^U observes x_{t-2}^U and m_{t-3}, this verifies our claim that the firm can infer m_{t-2} from x_{t-1}^U.

[15] Recall from Section I that "firms" are producer-consumers, and that minus the loss function is an approximation to an indirect utility function. Thus, finding the regime that minimizes the loss function is (approximately) the same as finding the one that maximizes utility of producer-consumers. Since the producer-consumers are the only agents in the model, it is natural to define the optimal regime as the one that maximizes their utility. (Although we refer to agents as firms, they are the economy's workers and consumers as well as its sellers; it is not the case that there are other agents whose utility we neglect.)

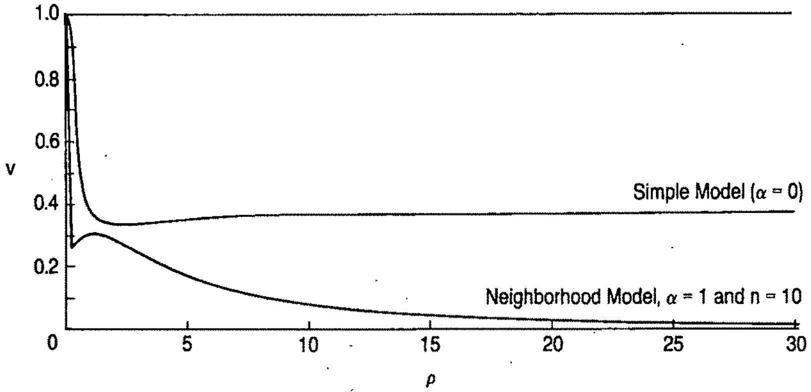


FIGURE 1. THE OPTIMAL TIMING OF PRICE CHANGES

greater than one, its value is unimportant in the simple model.[16]

C. *Results*: *The Neighborhood Model*

The results for the general model are complicated by the addition of two parameters: n, the size of a neighborhood; and $\alpha = \sigma_\theta / \sigma_\eta$, the relative importance of neighborhood and firm-specific shocks. (When $\alpha = 0$, the model reduces to the simple model.) For comparability with our earlier results, we fix n and α and graph the results in terms of v and ρ.[17] For $n = 10$ and $\alpha = 1$, the bottom line in Figure 1 shows the (v, ρ) locus for which $\mathscr{L}^S = \mathscr{L}^U$; staggering is optimal above the line. The results for other values of n and α are similar, and so we omit them.

Figure 1 shows that the welfare results for the neighborhood model are broadly similar to those for the simple model. One difference is that in the neighborhood model the value of ρ is important even for $\rho > 1$. In particular, as ρ approaches infinity, the $\mathscr{L}^S = \mathscr{L}^U$ locus approaches the horizontal axis, so staggering is optimal for all values of v. As ρ becomes large, monetary shocks disappear. In this case, the aggregate price level is stable, since it is unaffected by local shocks. Both of the costs of staggering—slow adjustment of the price level to monetary shocks, and changes in relative prices caused by changes in the price level when some prices are fixed—are eliminated. But staggering still has benefits: even without monetary shocks, firms want to distinguish neighborhood and firm-specific shocks, and thus value the information in prices set recently by neighbors.[18]

V. The Equilibrium Timing of Price Changes

While it is interesting to compare welfare under synchronization and staggering, it is also important to ask when each will arise in a decentralized economy. Therefore, we now assume that each firm chooses whether to change its price in even or odd periods. This allows us to determine when synchronization and staggering are stable Nash equilibria.

A. *Synchronization*

First consider synchronization. We study only the general model because the model without neighborhoods is similar. We assume that all firms set prices in even periods and ask whether a single "rebel" can gain by moving to odd periods. If not, then synchronization is a Nash equilibrium.

---

[16] According to Figure 1, the effect of ρ on the relative sizes of $\mathscr{L}^S$ and $\mathscr{L}^U$ is not monotonic. We have no explanation for this result.

[17] Recall that $\rho = \sigma_u / \sigma_m$. From (11), $\sigma_u^2 = 2\sigma_\theta^2 + \sigma_\eta^2$.

[18] As Lucas (1977) emphasizes, aggregate shocks are responsible for only a small part of the uncertainty facing firms in actual economies. Thus it seems realistic to assume that ρ is large, and hence that staggering is optimal for a wide range of v.

The rebel sets its price at t and $t+2$ (odd), while all other firms set prices at $t-1$ and $t+1$ (even). Since the economy is large, the rebel's behavior does not affect the aggregate price level. Only the rebel changes price at t, so $p_t^S = p_{t-1}^S$. Using (33),

$$p_t^S = p_{t-1}^S = m_{t-4} + \pi_1 \delta_{t-2} + \pi_2 \delta_{t-3} (t \text{ odd}). \tag{37}$$

All firms but the rebel change prices at $t+1$, so

$$p_{t+1}^S = m_{t-2} + \pi_1 \delta_t + \pi_2 \delta_{t-1} (t \text{ odd}). \tag{38}$$

Since $p_t \neq p_{t+1}$, the rebel uses the same price-setting rule as a firm in the staggered regime, (27). The rebel's information is similar, but not identical, to that of a staggered firm. Equation (37) implies that the rebel can infer m_{t-2} from prices set at $t-1$: p_{t-1}^S depends only on m_{t-2} and earlier money stocks, which have been announced. Since m_{t-2} is known, the rebel, like a firm under staggering, need only estimate δ_{t-1} and θ_{It-1}. And the rebel has the same pieces of information, $z_3^N = \delta_{t-1} + \theta_{It-1} - \theta_{It-2} + \eta_{it-1}$ and $z_4^N = \theta_{It-2} + \bar{\eta}_{it-2}$. But the results from projecting δ_{t-1} and θ_{It-1} on z_3^N and z_4^N are different from those in the staggered regime (see the Appendix). The reason is that $\bar{\eta}_{it-2}$ has a smaller variance. Under staggering, $\bar{\eta}_{it-2}$ is the average of the η's over firm i and the $(n/2)$ neighbors who set prices at $t-1$; thus the variance of $\bar{\eta}_{it-2}$ is $\sigma_\eta^2/(n/2+1)$. In contrast, since the rebel is the only firm that sets its price at t, it observes prices set by *all* its neighbors at $t-1$. Thus $\bar{\eta}_{it-2}$ is constructed from n observations and its variance is σ_η^2/n.

Calculations similar to the derivations of $\mathcal{L}^S$ and $\mathcal{L}^U$ lead to the rebel's loss, $\mathcal{L}^R$ (these calculations use the price-setting rule, (27), the equations for the price level, (37) and (38), and the appropriate projection formulas; again see the Appendix). Synchronization is an equilibrium if $\mathcal{L}^R > \mathcal{L}^S$—that is, if the rebel loses by breaking from synchronization.[19] For the simple model ($\alpha = 0$) and for $n = 10, \alpha = 1$, Figure 2 shows the (v, ρ) combinations for which $\mathcal{L}^R = \mathcal{L}^S$. Below each line, synchronization is an equilibrium; above the line, it is not.

The explanation for these results is that there are both costs and benefits to rebelling. The rebel gains information by observing prices set recently by all other firms. On the other hand, breaking from synchronization leads to large fluctuations in the rebel's relative price. If either v or ρ is small—so either relative price fluctuations have large costs or the information gains from staggering are small—no one breaks from synchronization. (Below, we compare the condition for synchronization to be an equilibrium to the condition for it to be socially optimal.)

B. *Uniform Staggering*

We now consider uniform staggering. The results for the simple model are very different from those for the neighborhood model.

The Simple Model. In the simple model, uniform staggering is clearly a Nash equilibrium for all parameter values. With no neighborhoods, the even- and odd-period cohorts of price setters are identical halves of the economy, and so the losses of firms in the two cohorts are the same. Further, since the economy is large, these losses do not change if a single firm switches cohorts. Thus no firm has an incentive to switch.

In the simple model, however, the staggered equilibrium is *never* stable. Stability of staggering is defined as follows: after a small perturbation in the cohort sizes away from half and half, firms in the larger group can gain by moving to the smaller one, restoring the equal sizes. To see whether this condition is met, we solve for the behavior of the economy when the cohorts have arbitrary

---

[19] More precisely, synchronization is an equilibrium if $\mathcal{L}^R \geq \mathcal{L}^S$, but it is a stable equilibrium only if $\mathcal{L}^R > \mathcal{L}^S$.

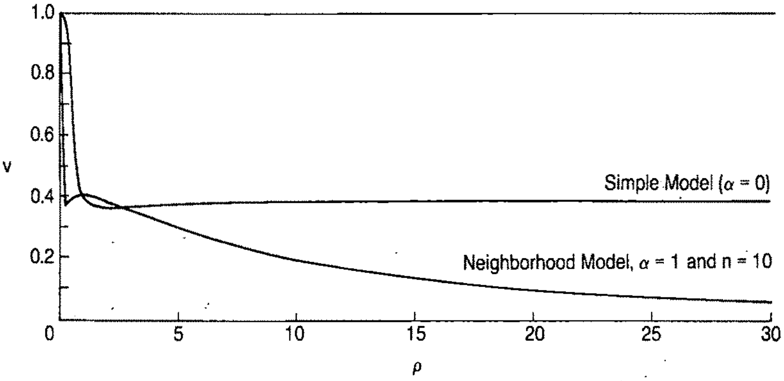


FIGURE 2. EQUILIBRIUM TIMING: SYNCHRONIZATION

sizes. (These calculations are similar in spirit to those for the special case of equal sizes; see the Appendix.)[20] We compute a firm's loss as a function of the proportion of firms in its cohort, k, and find that the derivative of this function evaluated at $k = \frac{1}{2}$ is always negative. Thus, following a perturbation away from uniform staggering, firms in the larger cohort are better off than firms in the smaller one. All firms want to join the larger group, and so uniform staggering is unstable.

This result arises because each firm wants to minimize fluctuations in its relative price, and therefore wants to synchronize its price setting with as many firms as possible. Crucially, this incentive to join the larger cohort is *not* offset by any information loss. To understand this, recall the information gain from uniform staggering in the simple model: a firm observes prices set at $t-1$ by half of all firms; each price reveals the sum of m_{t-2} and a local shock; and so the average of the prices reveals m_{t-2}. After a perturbation away from uniform staggering, firms in the large cohort observe prices set at $t-1$ by 50 minus percent of all firms (the small cohort), while firms in the small cohort observe prices set by 50 plus percent. But, since the economy is large, prices set by 50 plus percent and 50 minus percent of firms reveal the same information: in both cases, the local shocks affecting prices average to zero and m_{t-2} is revealed.

The Neighborhood Model. In the simple model, uniform staggering is unstable because each firm wants to join the larger of two price-setting cohorts. The importance of our assumptions about neighborhoods is that they give firms an incentive to join the smaller cohort. As a brief digression, note that there are other plausible modifications of the simple model that would do the same. For example, we could assume that gathering price data is costly. Suppose that the cost of observing enough prices set at $t-1$ to obtain a good estimate of m_{t-2} grows as the $t-1$ cohort shrinks. (A firm might need to travel farther to survey a given number of $t-1$ price setters.) In this case, each firm has an incentive to set its price *after* as many firms as possible—that is, to join the smaller cohort.[21]

---

[20] We derive price-setting rules for each cohort, aggregate to determine the behavior of the price level, and then combine these results to find the losses of firms in each cohort. The Appendix presents these calculations for the neighborhood model; the corresponding expressions for the simple model are obtained by setting α (which measures the importance of neighborhood shocks) to zero.

[21] Alternatively, one could assume that a firm observes only a small subset of other firms' prices. In this case, the firm obtains an imperfect estimate of x_{t-1}. When the prices observed by the firm include many set at $t-1$, the estimate of x_{t-1} is more precise, and therefore provides more information about m_{t-2}. This is an incentive for the firm to join the smaller cohort.

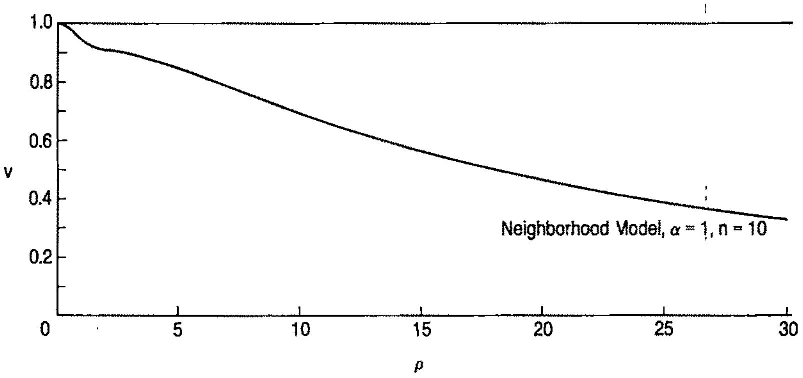


FIGURE 3. EQUILIBRIUM TIMING: STAGGERING

Returning to our neighborhood model, we now determine when uniform staggering is a stable equilibrium. As in the simple model, it is always an equilibrium, and so we focus on stability.[22] Recall that uniform staggering is defined as an equal split of each neighborhood into even- and odd-period price setters. We define stability as follows. Suppose that one firm in each neighborhood is moved from the odd to the even cohort, so that $((n/2)+1)$ firms in each neighborhood set prices in even periods and $((n/2)-1)$ set prices in odd periods. Staggering is stable if firms in the larger cohort have greater losses than firms in the smaller cohort.[23]

The Appendix computes the losses of firms in the even (larger) and odd (smaller) cohorts, $\mathscr{L}^E$ and $\mathscr{L}^O$. For $\alpha=1$ and $n=10$, Figure 3 shows the (v,ρ) combinations for which $\mathscr{L}^E=\mathscr{L}^O$. Staggering is stable above the line.

Staggering can be stable because firms in the smaller cohort have better information. A firm setting its price at t observes prices set at $t-1$. As described above, these reveal $\theta_{It-2}+\bar{\eta}_{it-2}$, where again $\bar{\eta}_i$ is the average of η over firm i and all neighbors setting prices at $t-1$. A firm in the smaller cohort observes more prices set at $t-1$, and therefore $\bar{\eta}_{it-2}$ has a smaller variance. $\theta_{It-2}+\bar{\eta}_{it-2}$ is a better estimate of θ_{It-2}, and so it is more useful in disentangling local and aggregate shocks. For some parameter values, this information advantage outweighs the disadvantage of changing prices at the same time as less than half the firms.

It is crucial that the neighborhoods are small. A firm that observes six neighboring prices set at $t-1$ obtains a better estimate of θ_{It-2} than a firm that observes four. In contrast, if the neighborhoods were large, a firm observing prices set by 50 plus percent of its neighborhood would learn no more than a firm observing 50 minus percent. The

---

[22] The equilibrium question is more subtle than in the simple model. Uniform staggering is a Nash equilibrium if no firm has an incentive to move from one of its neighborhood's cohorts to the other, given that all other neighborhoods remain equally divided. One can show that a firm that switches loses information because it synchronizes its price setting with more of its neighbors. There is no offsetting gain, because the firm still changes prices with 50 percent of the economy, and therefore experiences the same fluctuations in its relative price. No firm wants to switch, and so staggering is an equilibrium.

[23] There are other reasonable definitions of stability based on different perturbations. For example, we could move one firm in a single neighborhood from the odd to the even cohort while leaving the other neighborhoods unchanged and compare the losses of the two cohorts in the perturbed neighborhood. By this weak definition, the condition for stability is the same as the condition for staggering to be a Nash equilibrium, and it always holds (see fn. 22). Alternatively, we could perturb a small but nonnegligible proportion of the neighborhoods while leaving the others unchanged. With this definition, we conjecture that staggering is stable for a wider range of parameter values than in the text.

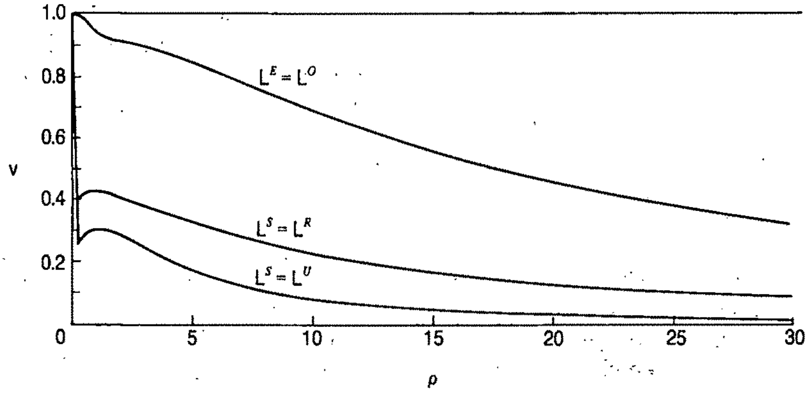


FIGURE 4. EQUILIBRIUM VS. OPTIMUM

η_{jt-2}'s would average to zero for both cohorts, and so all firms would observe θ_{It-2} exactly.

C. Comparison of Equilibrium and Optimum

The conditions under which synchronization and staggering are stable equilibria differ from the conditions under which they are socially optimal. This is shown by Figure 4, which summarizes the results in Figures 1–3 for the neighborhood model. In the region between the two lower lines in Figure 4, synchronization is an equilibrium even though staggering is optimal. In addition, between the top and bottom lines, staggering is optimal but not a stable equilibrium. According to these results, the incentives for an individual firm to break from synchronization or to join the smaller of two cohorts are weaker than the incentives for a social planner to choose staggering. Furthermore, for parameter values between the two upper lines, there is no stable equilibrium.

The relative positions of the three lines in Figure 4 do not appear robust. To take an example discussed above, suppose that gathering price data is costly. Let $c(k)$ be the cost of inferring m_{t-2} from prices set at $t-1$, where k is the proportion of the economy in a firm's cohort. Assume that $c'(k) > 0$: the cost of gathering price data rises as a firm's cohort grows and the $t-1$ cohort shrinks. To see the implications for the positions of the lines, recall that the bottom line is determined by comparing $\mathscr{L}^S$ to $\mathscr{L}^U$, and the middle line by comparing $\mathscr{L}^S$ to $\mathscr{L}^R$. This means that the relation between the two lower lines depends on $\mathscr{L}^U$ and $\mathscr{L}^R$. The cost of gathering prices under uniform staggering is $c(\frac{1}{2})$, while the cost to the rebel is $c(0)$, because no other firm belongs to the rebel's cohort. Since $c(\frac{1}{2})$ is greater than $c(0)$, introducing the cost raises $\mathscr{L}^U$ more than it raises $\mathscr{L}^R$. One can show that this raises the bottom line relative to the middle line, and that the two may switch positions. By a similar argument, the top line may move below the others. (The position of the top line depends on $c'(\frac{1}{2})$, which affects the benefit from belonging to the small cohort after a perturbation away from uniform staggering.)[24]

VI. Conclusion

Imperfect information can lead to staggered price setting. While fluctuations in relative prices are minimized when firms make decisions at the same times, a firm gains valuable information about demand if it waits to see other prices. This information gain can lead firms to break from synchronization and can make staggering a stable equilibrium. Thus staggered price adjustment, a crucial foundation of new Keynesian macroeconomic models, can arise from rational economic behavior.

---

[24] We also suspect that changing assumptions about the timing of announcements or the stochastic processes followed by shocks could change the relative positions of the lines.

Imperfect information can also make staggering socially optimal. Staggering leads to price-level inertia, which exacerbates aggregate fluctuations. However, by helping firms set prices closer to full-information levels, staggering creates efficiency gains that may outweigh the costs. Policy proposals to reduce staggering—for example, by requiring labor unions to sign contracts at the same times—could reduce welfare despite lessening inertia.

The model predicts that staggering is least likely when firms are nearly perfectly competitive and when idiosyncratic shocks are small. If the model is applied to an industry rather than the entire economy, these predictions are borne out by casual observation. In the automobile industry, products are fairly close substitutes and shocks are likely to affect all firms equally. The result is the synchronized pricing that we see. On the other hand, the drugstore on the corner and the diner next door produce goods that are poor substitutes, and they are likely to face different shocks. Thus it is not surprising that they change prices at different times.

This paper has studied a very specific model. Some results, notably the relation between the equilibrium and optimal regimes, are not likely to be robust. We doubt, however, that reasonable changes in the model would reverse the conclusion that staggering can be both optimal and a stable equilibrium if firms possess strong market power or idiosyncratic shocks are large.

A natural extension of our analysis is to alter the assumption that prices are fixed for two periods. If prices are fixed for longer, one can show that the incentive for a rebel to break from synchronization is greater. Thus the region of parameter values for which synchronization is an equilibrium is smaller. Intuitively, if prices are set for many periods, then a rebel setting its price one period after other firms is only slightly out of step. The cost in relative price fluctuations is small, while the rebel still gains the information in others' prices. Unfortunately, it is very difficult to determine how changing the frequency of price adjustment affects the socially optimal regime and the stability of uniform staggering.

Our intuition about the costs and benefits of staggering carries over to continuous time, but the model must be modified. In this paper, firms have perfect information about prices set in the most recent period. In a continuous time model, it would be unrealistic to assume that firms observe and respond to others' prices the instant they are set.[25] Instead, one might assume that firms learn about prices slowly, gaining full information after a discrete amount of time (for example, the time it takes to visit all neighboring stores). This modification of the model would complicate the analysis considerably.

APPENDIX

This appendix outlines the calculations omitted from the text. A more detailed version of the appendix is available from the authors. Throughout we consider the general model with neighborhood shocks. To obtain results for the simple model, set $\sigma_\theta^2 = 0$.

Projection Coefficients. Here we present the formulas for the projection coefficients in equations (23)–(25) and (30)—(31):

$$a_1^N = \frac{1+\rho_1^2+\rho_2^2}{D_1} \tag{A1}$$

$$a_2^N = \frac{-1-\rho_1^2}{D_1} \tag{A2}$$

$$a_3^N = \frac{\rho_2^2}{D_1} \tag{A3}$$

$$a_4^N = \frac{1+\rho_1^2+\rho_2^2}{D_1} \tag{A4}$$

$$a_5^N = \rho_1^2 a_1^N \tag{A5}$$

---

[25] For example, given the length of labor contract negotiations, one union's wage can influence another's only if the first union signs its contract significantly earlier.

$$a_6^N = \rho_1^2 a_2^N \tag{A6}$$

$$b_1^N = \frac{\rho_1^2 + \left(2\rho_2^2\right)/(n+2)}{D_2} \tag{A7}$$

$$b_2^N = \frac{-\rho_1^2}{D_2} \tag{A8}$$

$$b_3^N = \rho_1^2 b_1^N \tag{A9}$$

$$b_4^N = \rho_1^2 b_2^N, \tag{A10}$$

where

$$\rho_1 = \sigma_\theta/\sigma_m,\ \rho_2 = \sigma_\eta/\sigma_m,$$

$$D_1 = \left[2\left(1+\rho_1^2\right)+\rho_2^2\right]\left(1+\rho_1^2+\rho_2^2\right)-\left(1+\rho_2^2\right)^2,$$

and $D_2 = (1+2\rho_1^2+\rho_2^2)(\rho_1^2+2\rho_2^2/(n+2)-\rho_1^4$.

The Aggregate Price Level. Here we find the coefficients in the solutions for the price level under staggering and synchronization, (33) and (36). For synchronization, substitute (33) into (20) to obtain

$$\begin{aligned} x_{it}^S = m_{t-3} &+ [v+(1-v)\pi_1]E_{t-1}^i\delta_{t-1} \\ &+ [v+(1-v)\pi_2]E_{t-1}^i\delta_{t-2} + \frac{v}{2}E_{t-1}^i\theta_{It-1}. \end{aligned} \tag{A11}$$

Using (23)–(25), this becomes

$$\begin{aligned} x_{it}^S = m_{t-3} &+ \Big\{[v+(1-v)\pi_1]a_1^N \\ &\quad + [v+(1-v)\pi_2]a_3^N + \frac{v}{2}a_5^N\Big\}z_1^N \\ &+ \Big\{[v+(1-v)\pi_1]a_2^N \\ &\quad + [v+(1-v)\pi_2]a_4^N + \frac{v}{2}a_6^N\Big\}z_2^N. \end{aligned} \tag{A12}$$

Aggregating (which eliminates the local components of z_1^N and z_2^N) and matching coefficients with (33) yields the solutions for π_1 and π_2 in the text.

For staggering, substitute (34) and (35) into (27):

$$\begin{aligned} x_{it}^U = {} & \frac{(1-v)(1-\lambda_0)^2}{4}x_{t-1}^U \\ &+ \left[v + \frac{(1-v)(1-\lambda_0+2\lambda_1+\lambda_0\lambda_1)}{4}\right]E_{t-1}^i m_{t-1} \\ &+ \frac{(1-v)(2+\lambda_0)(1-\lambda_0-\lambda_1)}{4}m_{t-2} \\ &+ \frac{v}{2}E_{t-1}^i\theta_{It-1}. \end{aligned} \tag{A13}$$

Using (30)–(31), this leads to

$$\begin{aligned} x_{it}^U = {} & \frac{(1-v)(1-\lambda_0)^2}{4}x_{t-1}^U \\ &+ \left[v + \frac{(1-v)(\lambda_0+3)(1-\lambda_0)}{4}\right]m_{t-2} \\ &+ \left\{\left[v + \frac{(1-v)(1-\lambda_0+2\lambda_1+\lambda_0\lambda_1)}{4}\right]\right. \\ &\qquad \left. \times b_1^N + \frac{v}{2}b_3^N\right\}z_3^N \\ &+ \left\{\left[v + \frac{(1-v)(1-\lambda_0+2\lambda_1+\lambda_0\lambda_1)}{4}\right]\right. \\ &\qquad \left. \times b_2^N + \frac{v}{2}b_4^N\right\}z_4^N. \end{aligned} \tag{A14}$$

Aggregating and matching coefficients with (35) yields the solutions for λ_0 and λ_1 in the text.

Computation of $\mathscr{L}^S$ and $\mathscr{L}^U$. This section computes a firm's loss under synchronization, $\mathscr{L}^S$, and its loss under staggering, $\mathscr{L}^U$, to determine the optimal regime. To compute $\mathscr{L}^S$, recall that the coefficients on δ_{t-1} and δ_{t-2} in (A12) must equal π_1 and π_2. Thus (A12) can be rewritten as

$$\begin{aligned} x_{it}^S &= m_{t-3} + \pi_1 z_1^N + (\pi_2-\pi_1)z_2^N \\ &= p_t^S + \pi_1(\theta_{It-1}+\theta_{It-2}+\eta_{it-1}) \\ &\quad + (\pi_2-\pi_1)(\theta_{It-2}+\eta_{it-2}). \end{aligned} \tag{A15}$$

Combining (A15) with (33) and (6) yields $(x_{it}^S - p_{it}^*)$ and $(x_{it}^S - p_{it+1}^*)$ as functions of δ, θ, and η. Squaring these expressions and taking expectations yields

$$\begin{aligned} \mathscr{L}^S = {} & v^2\left[2(1-\pi_1)^2+2(1-\pi_2)^2+3\right]\sigma_m^2 \\ &+ \left[3v^2+(\pi_1-v)^2+\pi_1^2+2\pi_2^2\right]\sigma_\theta^2 \\ &+ 2\left[v^2+(\pi_2-\pi_1)^2+\pi_1^2\right]\sigma_\eta^2. \end{aligned} \tag{A16}$$

To compute $\mathscr{L}^U$, recall that the coefficients in (A14) must be consistent with the coefficients in (35). This leads to

$$\begin{aligned} x_{it}^U = x_t^U &+ \lambda_1(\theta_{It-1}+\theta_{It-2}+\eta_{it-1}) \\ &+ \left(b_2^N/b_1^N\right)\lambda_1(\theta_{It-2}+\eta_{it-2}). \end{aligned} \tag{A17}$$

Using (35), x_t^U can be rewritten as

$$x_t^U = \frac{(1-\lambda_0)m_{t-1}-(1-\lambda_0-\lambda_1)\delta_{t-1}}{(1-\lambda_0 B)}, \tag{A18}$$

where B is the lag operator. Substituting (A18) into (A17) and combining the result with (6) and (36) leads, after considerable algebra, to $(x_{it}^U - p_{it}^*)$ and $(x_{it}^U - p_{it+1}^*)$ as functions of the shocks. Squaring and taking expectations yields

$$\text{(A19)}\quad \mathscr{L}^U = \left[2v^2 + \beta_0 + \beta_2 + \frac{(\beta_0\lambda_0 + \beta_1)^2 + (\beta_2\lambda_0 + \beta_3)^2}{1-\lambda_0^2}\right]\sigma_m^2 + \left[3v^2 + \lambda_1^2 + (\lambda_1 - v)^2 + 2\lambda_1^2\left(1 + \left(b_2^N/b_1^N\right)\right)^2\right]\sigma_\theta^2 + 2\left[v^2 + \lambda_1^2 + \left(\frac{2}{n+2}\right)\left(\left(b_2^N/b_1^N\right)\lambda_1\right)^2\right]\sigma_\eta^2,$$

where

$$\beta_0 = \frac{\lambda_1(v+1) - 2v}{2}, \quad \beta_1 = \frac{(1-v)(1-\lambda_0-\lambda_1)}{2},$$

$$\beta_2 = \frac{\lambda_1(v-1) - 2v}{2} \quad \text{and} \quad \beta_3 = \frac{(1+v)(\lambda_1+\lambda_0-1)}{2}.$$

The Rebel. This section computes the rebel's price, x_{it}^R, and its projection formulas and uses these to find the rebel's loss, $\mathscr{L}^R$. Combining equations (8), (27), (37), and (38) leads to

$$\text{(A20)}\quad x_{it}^R = m_{t-2} + \frac{1-v}{2} \times \left[(\pi_1 - 1)\,\delta_{t-2} + (\pi_2 - 1)\,\delta_{t-3}\right] + \left[v + \frac{1-v}{2}\pi_2\right] E_{t-1}^i\,\delta_{t-1} + \frac{v}{2} E_{t-1}^i \theta_{It-1}.$$

As explained in the text, the rebel knows, m_{t-2} and, like a firm in the staggered regime, uses z_3^N and z_4^N to estimate δ_{t-1} and θ_{It-1}. But the variance of $\bar{\eta}_{it-2}$ is now σ_η^2/n. The rebel's projection equations are

$$\text{(A21)}\qquad E_{t-1}^i\,\delta_{t-1} = c_1^N z_3^N + c_2^N z_4^N$$

and

$$\text{(A22)}\qquad E_{t-1}^i \theta_{It-1} = c_3^N z_3^N + c_4^N z_4^N,$$

where the c^N's are formed from the b^N's defined above by replacing $(n/2)+1$ with n.

Substituting (A21) and (A22) into (A20) and combining the result with (6), (37), and (38) leads to $(x_{it}^R - p_{it}^*)$ and $(x_{it}^R - p_{it+1}^*)$. Squaring and taking expectations yields the rebel's loss

$$\text{(A23)}\quad \mathscr{L}^R = \left[2v^2 + 2\alpha_1^2 + 2\alpha_2^2 + \alpha_3^2 + (\alpha_0 - v)^2 + (\alpha_0 - \alpha_4)^2\right]\sigma_m^2 + \left[3v^2 + (\alpha_0 - v)^2 + \alpha_0^2 + 2(\alpha_0 + \alpha_6)^2\right]\sigma_\theta^2 + 2\left[v^2 + \alpha_0^2 + \frac{\alpha_6^2}{n}\right]\sigma_\eta^2,$$

where

$$\alpha_0 = Ac_1^N, \quad \alpha_1 = \frac{1-v}{2}(\pi_1 - 1),$$

$$\alpha_2 = \frac{1-v}{2}(\pi_2 - 1), \quad \alpha_3 = v + (1-v)\pi_1,$$

$$\alpha_4 = v + (1-v)\pi_2, \quad \alpha_5 = \alpha_0 - \alpha_4,$$

$$\alpha_6 = Ac_2^N, \quad \text{and} \quad A = \left[v + \frac{1-v}{2}\pi_2 + \frac{\rho_1^2 v}{2}\right].$$

Is Uniform Staggering Stable? This section studies the neighborhood model when an arbitrary proportion of each neighborhood, k, belongs to the even cohort. We use the results to determine when uniform staggering is a stable equilibrium.

Let p_t^E be the price level at an even t, and let p_{t+1}^O be the price level at $t+1$ (odd). From the definition of the aggregate price level,

$$\text{(A24)}\qquad p_t^E = kx_t^E + (1-k)x_{t-1}^O; \quad \text{and}$$

$$p_{t+1}^O = (1-k)x_{t+1}^O + kx_t^E.$$

Combining (A24) with (6) and (8) yields price-setting rules for firms in the even and odd cohorts

$$\text{(A25)}\qquad x_{it}^E = vE_{t-1}^i m_{t-1} + \frac{v}{2}E_{t-1}^i\theta_{It-1} + \frac{1-v}{2}\left[(1-k)E_{t-1}^i x_{t+1}^O + 2kE_{t-1}^i x_t^E + (1-k)x_{t-1}^O\right]$$

and

$$\text{(A26)}\quad x_{it+1}^O = vE_t^i m_t + \frac{v}{2}E_t^i\theta_{It} + \frac{1-v}{2}\left[kE_t^i x_{t+2}^E + 2(1-k)E_t^i x_{t+1}^O + kx_t^E\right].$$

Firms in both the even and odd cohorts use z_3^N and z_4^N to estimate δ_{t-1} and θ_{It-1}. However, the projection equations for the two cohorts differ because $\bar{\eta}_{it-2}$ has different variances: $\mathrm{Var}(\bar{\eta}_{it-2})=\sigma_\eta^2/[(1-k)n+1]$ for the even cohort and $\mathrm{Var}(\bar{\eta}_{it-2})=\sigma_\eta^2/(kn+1)$ for the odd cohort. The estimates of firms in the even cohort are given by (30) and (31) with the b^N's replaced by b^E's, which are formed by replacing $(n/2)+1$ with $(1-k)n+1$ in (A7)–(A10). The estimates for the odd cohort are given by the estimates for the even cohort with k replaced by $1-k$.

The price-setting rules (A21) and (A22) imply solutions for x_t^E and x_{t+1}^O of the form

$$\text{(A27)}\quad x_t^E=\lambda_0^E x_{t-1}^O+\lambda_1^E m_{t-1}+\left(1-\lambda_0^E-\lambda_1^E\right)m_{t-2}$$

and

$$\text{(A28)}\quad x_{t+1}^O=\lambda_0^O x_t^E+\lambda_1^O m_t+\left(1-\lambda_0^O-\lambda_1^O\right)m_{t-1}.$$

To solve for the λ's, first use the projection equations to find the expectation of (A27). Use this result, and the projection equations again, to find the expectation of (A28). Substitute the two results into (A25), aggregate, and set the coefficients on x_{t-1}^O and m_{t-1} equal to λ_0^E and λ_1^E. Symmetry yields similar expressions for λ_0^O and λ_1^O. Finally, combining these four expressions yields

$$\text{(A29)}\quad \lambda_0^E=\frac{(v+k-vk)}{k(1-v)}-\frac{\sqrt{(v+k-vk)^2-\left[(v+k-vk)(1-k)k(1-v)^2\right]/(1-k+vk)}}{k(1-v)},$$

and

$$\text{(A30)}\quad \lambda_1^E=\frac{b_1^E\left[v(2+\rho_1)-(1-v)(1-k)\left(1-\lambda_0^O\right)\right]}{2-b_1^E(1-v)\left(2k+(1-k)\lambda_0^O\right)}.$$

The solutions for λ_0^O and λ_1^O are obtained by replacing k with $1-k$ in (A29) and (A30).

By reasoning analogous to the derivation of (A17), we can write x_{it}^E as

$$\text{(A31)}\quad x_{it}^E=x_t^E+\lambda_1^E(\theta_{It-1}+\theta_{It-2}+\eta_{it-1})+\left(b_2^E/b_1^E\right)\lambda_1^E(\theta_{It-2}+\bar{\eta}_{it-2}).$$

Combining this with (6) and (A24) leads to $(x_{it}^E-p_{it}^*)$ and $(x_{it}^E-p_{it+1}^*)$. Squaring and taking expectations yields the loss of a firm in the even cohort:

$$\text{(A32)}\quad \mathscr{L}^E=\left\{2v^2+\gamma_0^2+\gamma_3^2+\frac{\gamma_1^2+\gamma_4^2+\left(\gamma_0\lambda_0^E\lambda_0^O+\gamma_2\right)^2+\left(\gamma_3\lambda_0^E\lambda_0^O+\gamma_5\right)^2}{1-\left(\lambda_0^E\lambda_0^O\right)^2}\right\}\sigma_m^2+\left\{3v^2+\left(\lambda_1^E-v\right)^2+\left(\lambda_1^E\right)^2+2\left[\lambda_1^E\left(1+\frac{b_2^E}{b_1^E}\right)\right]^2\right\}\sigma_\theta^2+2\left\{v^2+\left(\lambda_1^E\right)^2+\left[\frac{1}{(1-k)n+1}\right]\left[\lambda_1^E\frac{b_2^E}{b_1^E}\right]^2\right\}\sigma_\eta^2,$$

where

$$\gamma_0=(1-k+vk)\lambda_1^E-v,$$

$$\gamma_1=\left(1-\lambda_1^O\right)\left[(1-k+vk)\left(1-\lambda_0^E\right)-v\right],$$

$$\gamma_2=(1-v)(1-k)\lambda_0^O\left(1-\lambda_0^E-\lambda_1^E\right],$$

$$\gamma_3=-\left[v+(1-v)(1-k)\lambda_1^O\right],$$

$$\gamma_4=\gamma_0-(1-v)(1-k)\left(1-\lambda_0^O+\lambda_1^E\lambda_0^O\right),$$

and

$$\gamma_5=\gamma_1-(1-v)(1-k)\left(1-\lambda_0^O-\lambda_1^E\lambda_0^O\right)+v\lambda_0^E\lambda_0^O.$$

To find the loss of a firm in the odd cohort, replace $(1-k)$ with k, replace the b^E's with the corresponding coefficients for the odd cohort, and switch all λ^E's and λ^O's. As explained in the text, uniform staggering is stable if $\mathscr{L}^E>\mathscr{L}^O$ when $(n/2)+1$ firms in each neighborhood belong to the even cohort (that is, when $k=\frac{1}{2}+(1/n)$).

REFERENCES

Ball, Laurence, "Externalities from Contract Length," *American Economic Review*, September 1987, *77*, 615–629.

______ **and Romer, David,** (1987a) "Are Prices Too Sticky?," NBER Working Paper No. 2171, January 1987.

______ **and** ______, (1987b) "The Equilibrium and Optimal Timing of Price Changes," NBER Working Paper No. 2412, October 1987.

Blanchard, Olivier J., "Price Asynchronization and Price Level Inertia," in R. Dornbusch and M. Simonsen, eds., *Inflation, Debt and Indexation*, Cambridge: MIT Press, 1983, 3–24.

______, "The Wage-Price Spiral," *Quarterly Journal of Economics*, August 1986, *101*, 543–65.

______ **and Kiyotaki, Nobuhiro,** "Monopolistic Competition and the Effects of Aggregate Demand," *American Economic Review*, September 1987, *77*, 647–66.

______ **and Fischer, Stanley,** *Macroeconomics*, mimeo., Cambridge: MIT, 1988, ch. 9.

Fethke, Gary and Policano, Andrew, "Wage Contingencies, the Pattern of Negotiation and Aggregate Implications of Alternative Contract Structures," *Journal of Monetary Economics*, September 1984, *14*, 151–70.

______ **and** ______, "Will Wage Setters Ever Stagger Decisions?," *Quarterly Journal of Economics*, November 1986, *101*, 867–77.

Gray, Jo Anna, "Wage Indexation: A Macroeconomic Approach," *Journal of Monetary Economics*, April 1976, *3*, 221–35.

Hall, Robert E. and Taylor, John B., *Macroeconomics*, New York: W. W. Norton, 2nd ed., 1988.

Lucas, Robert E., Jr., "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review*, June 1973, *63*, 326–34.

______, "Understanding Business Cycles," in K. Brunner and A. Meltzer, eds., *Stabilization of the Domestic and International Economy*, Carnegie-Rochester Conference on Public Policy, 5: 1977, 7–29.

Maskin, Eric S. and Tirole, Jean, "Models of Dynamic Oligopoly II: Competition through Prices," MIT Department of Economics Working Paper No. 373, 1985.

Okun, Arthur M., *Prices and Quantities*, Washington, D.C.: The Brookings Institution, 1981.

Parkin, Michael, "The Output-Inflation Trade-Off When Prices Are Costly to Change," *Journal of Political Economy*, February 1986, *94*, 200–24.

Taylor, John B., "Aggregate Dynamics and Staggered Contracts," *Journal of Political Economy*, February 1980, *88*, 1–23.

The Marginal Efficiency Cost of Redistribution

By CHARLES L. BALLARD*

The efficiency effects of redistributive policies are analyzed using a computational general equilibrium model. For parameter values considered most reasonable, a tax-financed universal cash grant generates losses for the higher-income groups that exceed the gains of the lower-income groups by from 50 to 130 percent. These efficiency costs are substantially below those calculated by other authors. The marginal efficiency cost of redistribution through wage subsidies is near zero, and can be negative.

Economists have long recognized that there may be a tradeoff between equality and efficiency. If transfers and subsidies are used to improve the economic condition of the poor population by one dollar, the distortions may be such that the cost to other members of society is more than one dollar. The excess loss is called the marginal efficiency cost of redistribution (MECR). The size of the MECR is critically important to our evaluation of programs designed to reduce poverty and inequality.

In this paper, I present a simple computational general equilibrium (CGE) model, with which I evaluate the MECR of various redistributive programs under a variety of assumptions about tax rates and other parameter values. The emphasis in this static model is on the effects of taxes and transfers on labor supply. Two main categories of redistributive policy change are considered here. The first type involves increases in the cash transfers; the second involves increases in net wages for the low-income population, which I will refer to as wage subsidies. In every case, tax rates are increased in order to finance the redistributive program.

For the combinations of parameter values that I consider most reasonable, a universal cash grant program financed with taxes on all consumer groups involves marginal efficiency costs in the range of 50¢ to $1.30 per dollar of redistribution. In other words, when the lower-income groups are made better off by a dollar, the higher-income groups are made worse off by $1.50 to $2.30. These welfare costs are substantial, but they fall well below those reported by Edgar Browning and William Johnson (1984), even though the elasticity and tax rate parameters used are very similar.

The MECRs for wage subsidy schemes tend to be close to zero and are often negative.[1] I also calculate "social welfare" for additive social welfare functions. A parameter of the social welfare function determines the degree of society's aversion to inequality. I find that, even when the MECRs are high, social preferences do not need to be anywhere near to Rawlsian for society to prefer any of these redistributive policy changes.

---

*Department of Economics, Michigan State University, East Lansing, Michigan 48824. I am grateful to Edgar Browning, John Goddeeris, Roger Gordon, Larry Goulder, Yolanda Henderson, Harry Holzer, Bill Johnson, Ken Judd, Larry Martin, Paul Menchik, Karl Scholz, John Shoven, Jon Skinner, participants in seminars at the NBER Summer Institute and at Michigan State University, and four anonymous referees for helpful comments. I am grateful to the National Science Foundation under grant no. SES-8520283 for financial support. I alone am responsible for any errors. I thank Maxine Schafer and Terie Snyder for their fine work in preparing the manuscript, and Charles Chesbrough, Edward Guzzo, Nader Habibi, and Steven Medema for fine research assistance.

[1] Other authors have studied wage taxes, wage subsidies, etc., in a variety of types of models, and similar conclusions have been reached in many cases. Among the works in this literature are Alan Blinder, 1974; Sören Blomquist, 1981, 1984; Edgar Browning, 1973; Jonathan Kesselman, 1976a, 1977; Peter Wagstaff, 1975; and Richard Zeckhauser, 1971.

I. Model and Methodology

The model includes 384 consumer groups, which are divided into 10 household types. The data will be described in the next section. For now, it should be said that most of the household types have only a single earner, but one of the types is married couples where both spouses work. In the two-earner case, it is necessary to have a different utility structure from that of the one-earner case. In describing the specification of the utility functions and welfare evaluations, I will present the equation that applies to the one-earner households. Generalization to the two-earner households is straightforward.

Each one-earner household maximizes a constant elasticity of substitution (CES) utility function defined over leisure, l, and consumption of goods, X. Thus, each household's utility function takes the form

$$(1) \qquad U=\left[(1-\beta)^{1/\varepsilon}X^{(\varepsilon-1)/\varepsilon} + \beta^{1/\varepsilon}l^{(\varepsilon-1)/\varepsilon}\right]^{\varepsilon/(\varepsilon-1)},$$

where ε is the elasticity of substitution and β is a weighting parameter. For simplicity, I have suppressed the subscript that would indicate which consumer is being considered.

The budget constraint states that the net-of-tax value of the household's endowments of capital and labor, plus transfers received, should equal expenditure on goods and leisure. Thus, for the one-earner households,

$$(2) \qquad P_lE+P_KK(1-\tau_K)+\mathrm{TR}=P_ll+PX,$$

where P_l is the price of leisure or the net wage, P_K is the gross price of capital services, τ_K is the capital tax rate, E is the consumer's leisure endowment, TR is transfers received, and P is the price of consumption goods. P_l is the gross wage, P_L, less the tax on labor income at the rate of τ_L. Note that the consumer has an endowment of only one type of labor. The paper abstracts from issues of occupational heterogeneity.

Constrained maximization leads to the demand functions

$$(3) \qquad X=\frac{(1-\beta)I}{P^{\varepsilon}\Delta}$$

$$(4) \qquad l=\frac{\beta I}{P_l^{\varepsilon}\Delta},$$

where I is the net value of endowments plus transfers, and

$$(5) \qquad \Delta=(1-\beta)P^{1-\varepsilon}+\beta P_l^{1-\varepsilon}.$$

The choices of the utility function parameters are discussed in the next section. There is only one production sector in this model. Technology is described by a CES production function, which takes the form

$$(6) \qquad X=\phi\left[\alpha L^{(\sigma-1)/\sigma} + (1-\alpha)K^{(\sigma-1)/\sigma}\right]^{\sigma/(\sigma-1)},$$

where X is output, L and K are the inputs of labor services and capital services, σ is the elasticity of substitution in production, α is a weighting parameter, and ϕ is a scale parameter. The production sector minimizes the cost of producing the output that is demanded. The choices of the production function parameters are discussed in the next section.

The government in this model collects taxes on the labor incomes and capital incomes of all consumer groups.[2] The revenues are used to make transfer payments and exhaustive expenditures. The transfer payments enter the consumer's budget constraint, and thus can be "spent" either on goods or on leisure. In the case of exhaustive expenditures, the government merely purchases and consumes some of the output of the production sector. In this model, exhaustive government expenditure and private consumption are "ordinary independents" (in the words of David Wildasin, 1984).

---

[2] In a static model such as this, capital taxes are nondistortionary. I include them only in order to give slightly more realism to the parameterizations.

The analysis proceeds in the following manner. Once the data are collected, a set of tax and elasticity parameters is chosen. The utility and production function parameters are then chosen using the technique of "calibration," which will be discussed in the next section. When the model has been calibrated, a "base case" simulation is run, and information on base case prices and quantities is saved. Next, a series of "revised case" simulations is performed using the same utility and production function parameters but different configurations of tax, government spending, and transfer parameters. I use a modified version of the tâtonnement algorithm suggested by Larry Kimbell and Glenn Harrison (1984) to solve for equilibrium. Finally, welfare comparisons are made by comparing revised case prices and quantities with base case prices and quantities.[3]

The welfare numbers reported in Section III are based on equivalent variations. The equivalent variation is preferred to the compensating variation when we desire to compare several different policy changes, since the equivalent variation uses the same set of prices for every evaluation. The compensating variation, on the other hand, is measured in terms of the new prices that will be different for each new simulated equilibrium. However, since the changes between the base case and the revised case in this paper are small, the compensating and equivalent variations are very close to each other. For example, for my central case simulation of a universal cash grant plan, reported in Section III, the MECR based on the compensating variation is 80.111¢, while the MECR based on the equivalent variation is 80.114¢.

In order to allow for fair comparisons across different types of policy changes, it is necessary to hold constant the amount of marginal tax increase. In virtually all of the simulations in this paper, I raise tax rates by one-tenth of one percentage point. In this way, I control for the budgetary size of the marginal policy change. I also perform some sensitivity analysis with respect to the size of the marginal tax increase. Larger marginal changes lead to slightly larger welfare costs.

I evaluate several types of redistributive policy changes. The first of these is one under which the cash grant is increased for all groups, and the marginal tax rate is also increased for all groups. I call this a universal tax-financed cash grant. I consider two types of universal cash grant. In the first, the increase in the cash grants is assumed to be the same for every person. This type of policy has been called a "demogrant." The second case involves increased cash grants that are the same for every household.[4] The results from these policy changes differ because the number of persons per household is not constant across income classes. It should be noted that neither of these sets of additional transfer payments is proportional to the existing transfer payment system. The existing system consists of many different programs, each with its own eligibility requirements. Most actual cash grant transfer programs fall between the two extremes of equal-per-capita and equal-per-household grants. Thus, the two formats may be seen as representing convenient upper and lower bounds for the marginal efficiency costs of universal cash grant programs.

I also consider a "notch" cash grant program, under which cash grants are increased only for the lowest-income groups, and taxes are increased only for the highest-income groups. (I allocate the transfers on an equal per person basis.) In practice, a notch could have severe adverse efficiency effects. I do not directly assess the efficiency cost of the program with the notch, since for any given consumer I maintain a linear budget constraint. Nevertheless, the results are interesting because they indicate that the notch program may be more efficient than the universal cash grant program, if the notch can be placed in an area of the income distribution where its negative effects are minimized.

---

[3] More details on the welfare evaluation procedures can be found in an appendix, available upon request from the author. In addition, the author can provide more details on the construction of the data set and on calibration procedures, as well as on a wide range of sensitivity analyses beyond those reported here.

[4] Note that I consider household structure to be exogenous. This assumption may impart a downward bias to my calculations of the welfare cost of equal-per-household cash grants, since such grants may actually induce some reorganization of households.

Another policy considered here is a "wage subsidy," which is modeled as an increase in the net-of-tax wage rate for some of the low-income groups. This can be financed in a number of ways, but I most often consider increases in the marginal income tax rates for all consumer groups. I consider a variety of wage subsidy plans. In some cases, the subsidy applies to all workers in the lowest income strata. In other cases, it does not apply to the primary earner in two-earner households. Since a wage subsidy is of little use to a household with no earnings potential, I also consider wage subsidy plans which are supplemented with cash grants for the nonworking groups.

The stylized policy proposals considered here are considerably less complex than most proposals debated before Congress. However, most "real world" proposals share some of the characteristics of the proposals in this paper. Any policy proposed that involves increasing the net-of-tax wage rates of low-income workers will have some of the favorable efficiency effects of the simple "wage subsidy" that I simulate here. This will be true regardless of the name given to the real world policy proposal. For example, an important element of the early negative income tax proposals was a reduction in the implicit marginal tax rate faced by welfare recipients. My simulations suggest that this type of proposal would have relatively good efficiency properties. The same may be true of public employment projects where workers are paid at rates exceeding their market wages, although the exact value of these projects would depend on the type of output that is produced.

II. Data, Parameter Selection, and Model Specification

The simulations reported in this paper employ a data set consisting of 384 household groups. These groups represent the United States population for 1979. The data are taken from the 1980 Census of Population (U.S. Department of Commerce, 1984) and from Current Population Reports (U.S. Department of Commerce, 1981, 1983). The population is disaggregated along three dimensions. First, there are ten different household types, which are distinguished according to marital status, gender of head, and labor force status. Second, there are eight income classes. Third, the household/income cells are further broken down by number of persons per household. The classifications by household type and by income stratum are shown in Table 1, along with some other information for these groups. The complete data set contains information on the incomes from labor, capital, and transfers for each household group.

The choice of labor tax rates is complicated by the fact that individual income tax rates have decreased in recent years, so that no single simulation can be fully applicable to all of the recent experience in the United States. The appropriate tax rate for the simulations would include the combined effects of payroll taxes as well as income taxes, both at the state and federal level. For the central case, I choose 40 percent as the marginal labor tax rate, and 40 percent as the capital tax rate. As a stylized starting point, it is assumed that these rates apply to all consumer groups. For sensitivity analysis, I also consider the cases where different assumed levels of initial government purchases or transfers lead to labor tax rates varying from as low as 30 percent to as high as 60 percent (for all consumers). In addition, I consider initial capital tax rates ranging from zero up to 60 percent. Differences in the initial labor tax rates have a substantial effect on the marginal welfare costs. Differences in the initial capital tax rates have very little effect on the results, since capital taxes have no first-order distortionary effect in this model. I also provide sensitivity analysis with respect to the assumption that all groups face the same tax rates. I investigate cases in which the marginal labor tax rate faced by the lowest-income groups ranges from 25 percent to 60 percent.

TABLE 1—CLASSIFICATION OF HOUSEHOLD TYPES AND INCOME LEVELS IN THE 1979 DATA SET, ALONG WITH OTHER DATA

Household Type	Percentage of all Households	Average Number of Persons Per Household
1. Married Couples, Both Work	29.99 percent	3.06
2. Married Couples, Husband Works	30.89	2.93
3. Male-Headed Families, No Wife Present	2.48	2.38
4. Female-Headed Families, No Husband Present, Woman Works	6.13	2.70
5. Female-Headed Families, No Husband Present, Woman Does Not Work	4.07	2.51
6. Male Householders Living Alone	8.79	1.00
7. Female Householders Living Alone, Woman Works	6.80	1.00
8. Female Householders Living Alone, Woman Does Not Work	7.03	1.00
9. Male-Headed Households with Two or More Persons	2.32	2.31
10. Female-Headed Households with Two or More Persons	1.50	2.31

Income Class (in 1979 Dollars)	Percentage of All Households	Average Number of Persons per Household	Average Income per Household
1. $0–4,999	13.46	1.76	$3,481
2. $5,000–9,999	16.05	2.08	8,440
3. $10,000–14,999	15.40	2.35	13,396
4. $15,000–19,999	14.11	2.61	17,694
5. $20,000–24,999	12.37	2.79	22,238
6. $25,000–34,999	15.58	2.88	28,869
7. $35,000–49,999	8.52	2.84	40,548
8. $50,000 and over	4.52	2.80	73,069

Sources: United States Department of Commerce, Bureau of the Census (1981, 1983, 1984).

The labor supply elasticities for the various consumer groups are very important for the results of the model. Extensive reviews of the econometric literature on labor supply estimation can be found in James Heckman, Mark Killingsworth, and Thomas MaCurdy (1981) and in Killingsworth (1983). The studies emanating from the four negative income tax experiments are reviewed by Michael Keeley (1981) and Gary Burtless (1987). The various strands of the literature seem to agree that the labor supply responses of males are rather inelastic (and often slightly backward-bending), and that female labor supply is positively elastic. Unfortunately, there is wide variation in the magnitudes of the estimates of female labor supply responses. I consider a wide variety of combinations of labor supply elasticities. The one case on which I focus the most has an uncompensated elasticity of 1.0 for women, −0.05 for men, and 0.20 for married couples. These give a weighted average elasticity of about 0.16 for the economy. The weighted average of the compensated elasticities is about 0.36. Extensive sensitivity analysis is provided.

For the elasticity of substitution in production, I refer the reader to Ernst Berndt (1976) or Daniel Hamermesh (1986). In my simulations, I use an elasticity of 0.8 most often, but I also consider elasticities from 0.6

to 20.0, providing an extremely wide range of sensitivity analysis.

Once values have been selected for these key elasticity parameters, we can solve for the utility function and production function parameters that are consistent with the data. This procedure of "calibrating" certain parameter values is used widely with CGE models. In brief, ε (the consumer's elasticity parameter) is calibrated by using the assumed labor supply elasticity and manipulating the leisure demand equation. Once we have solved for ε, the consumer's first-order conditions are used to solve for β, the leisure-weight parameter. Given a value of σ, the elasticity of substitution in production, and the first-order conditions for cost minimization, we can solve for α, the production function weight parameter. Once α is known, the zero-profit condition is used to solve for ϕ, the scale parameter.

III. Simulation Results

We begin with the case in which each of the household groups has the same labor supply elasticity, even though this is clearly not the most realistic assumption. The first set of simulations involves initial factor tax rates of 40 percent, an uncompensated labor supply elasticity of 0.15 for all groups, and a compensated elasticity of 0.35 for all groups. To calculate the MECR under a universal cash grant plan, we increase the transfers and tax rates facing all groups. The extra transfers are allocated equally on a per capita basis. The MECR of such a plan is 71.8¢. The MECR for a cash grant plan with a notch is 14.3¢.[5]

An intuitive understanding of the wide difference between the MECRs for these two plans can be gained if we imagine that a tracking device could be attached to every dollar as it flows into and out of the Treasury. To the extent that some dollars are taxed away from the high-income groups to finance transfers for the low-income groups, there is an element of pure redistribution. This is indeed what occurs in the case of the notch program. However, under the universal tax-financed cash grant, some of the taxes raised from each group could be viewed as being paid back to the same group. This is pure waste. When a dollar is raised from an individual with distortionary taxes and paid back to the same individual in a lump sum, there is a welfare loss because of the substitution effect. Finally at least some of the dollars raised in taxes from the poor under the universal cash grant could be viewed as being paid to the rich. This obviously is not an efficient way to redistribute from rich to poor.

It is interesting that the difference between the MECRs for the notch cash grant and the universal cash grant is so large. This suggests that, even in a model in which the efficiency effects of the notch were recognized explicitly, the notch plan still might be a superior means of redistribution. In fact, the notches that we observe in some transfer payment programs in the United States are at income levels that may affect very large numbers of people, so that the welfare effects may be very great. Nevertheless we cannot rule out the possibility that the notch program would be superior. Explicit consideration of the welfare effects of the notches would be an interesting extension for future work. For a recent discussion of notches and their effects, see Blinder and Harvey Rosen (1985).

Finally, let us consider the MECR for a wage subsidy plan, when the poorest two income classes are the recipients of the wage subsidies. Here, the MECR is 1.2¢, or about 2 percent of the MECR for the universal tax-financed cash grant. Why is the MECR for the wage subsidy so close to zero? When a wage subsidy is applied to low-income groups, it offsets the positive marginal tax rates that they already face. However, in order to pay for the wage subsidy, higher taxes must be assessed. (In these simulations,

---

[5] For purposes of comparison, the marginal welfare cost (MWC) of additional government exhaustive expenditure is 10.3¢ when these same parameters are used. For the various parameter combinations considered here, the MWC is between 5¢ and 30¢, per dollar of additional exhaustive expenditure. See Ballard (1987c) for a comparison of MWC and MECR calculations. See also Ballard (1987a) and Ingemar Hansson and Charles Stuart (1987).

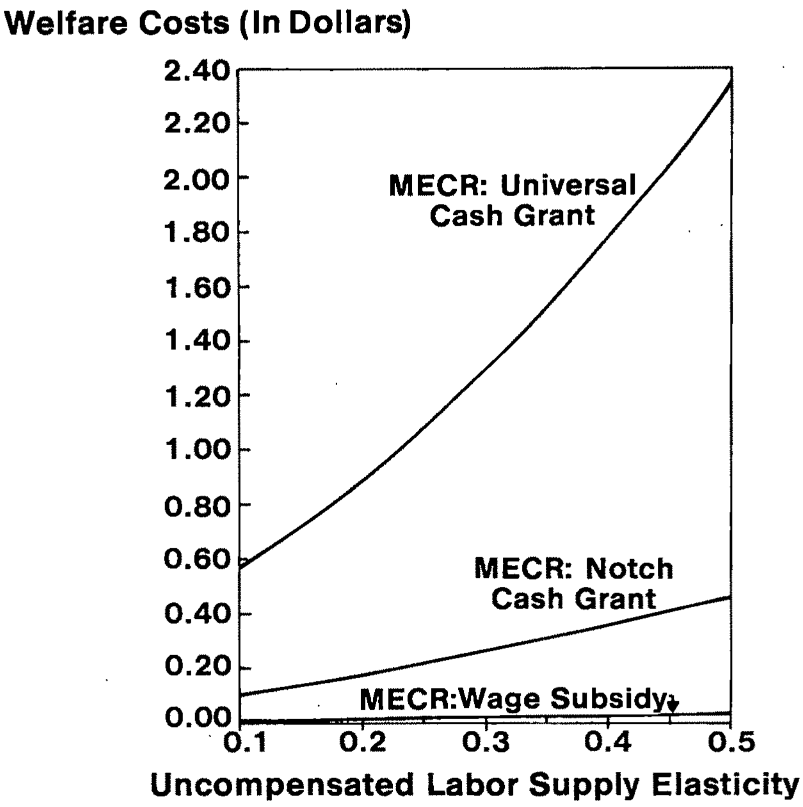


FIGURE 1. MARGINAL EFFICIENCY COST OF REDISTRIBUTION AS A FUNCTION OF THE UNCOMPENSATED LABOR SUPPLY ELASTICITY. ALL GROUPS HAVE THE SAME ELASTICITIES IN EACH CASE

the extra taxes are collected from all groups.) Thus, the wage subsidy reduces one distortion and increases another. When all groups are assumed to have the same labor supply elasticity and the same inital tax rate, these changes will be virtually identical. The gain from reducing one tax rate to slightly below 40 percent is nearly as great as the loss from increasing one tax rate to slightly above 40 percent.

In Figure 1, I show the sensitivity of these results to changes in the assumed uncompensated labor supply elasticity. Clearly, the results depend importantly on the labor supply elasticity. The greater one believes the elasticity to be, the worse cash grants will be when compared with the wage subsidy. Even with an elasticity of 0.5 (which I consider quite high), the MECR is still only 3.6¢ per dollar of redistribution for the wage subsidy, whereas it is $2.32 for a universal cash grant.

The assumption that all groups have the same elasticity was convenient for expositional purposes, but it was clearly unrealistic. I now develop my "central case" assumptions, in which different groups are characterized by different labor supply elasticities. I assume that the households in which the only earner is male have an uncompensated elasticity of -0.05. The female-headed households have an uncompensated elasticity of 1.0. Married couple households have a weighted average elasticity of 0.20, with the wife's own-wage elasticity equaling 1.0. These values yield an overall weighted average for the entire economy of 0.161. I assume that the total income elasticity of

TABLE 2—MARGINAL EFFICIENCY COST OF REDISTRIBUTION AND MARGINAL WELFARE COST AS A FUNCTION OF LABOR SUPPLY ELASTICITIES IN THE ECONOMY. (FOR EVERY GROUP, TOTAL INCOME ELASTICITIES ARE 0.2)

					MECR		
Case	Men's Elasticity	Elasticity for Married Couples	Women's Elasticity	Weighted Average Elasticity for the Economy	Universal Tax-Financed Cash Grant	Notch Cash Grant	Wage Subsidy
1	−0.10	0.10	0.75	0.070	51.9¢	10.5¢	−6.6¢
2 (Central Case)	−0.05	0.20	1.00	0.161	80.1	17.7	−8.0
3	0.00	0.30	1.25	0.253	114.4	25.3	−9.4
4	0.05	0.40	1.50	0.344	155.6	33.3	−10.7
5	0.10	0.50	1.75	0.435	203.0	45.9	−12.0

labor supply is −0.2 for every group, so that the weighted average compensated elasticity is 0.361. Once, again, the factor tax rates are set to 40 percent.

The results for this central case set of elasticities are similar to those reported above. The MECR for a notch cash grant is in the same vicinity as before. If the recipients of transfers under the notch are the bottom-income class only, the MECR is 14.3¢. If the bottom two classes both receive transfers, the MECR is 17.7¢. The MECR for the universal tax-financed cash grant is also in a similar range as before. It takes on a value of 80.1¢ when the transfers are allocated on a per capita basis. If the transfers are allocated on a per household basis, the MECR is 70.0¢. The efficiency cost of the per household transfers is less because the poorer income classes have fewer persons per household. Consequently, when transfers are allocated equally to every household, they are more precisely targeted.

The wage subsidy MECRs are slightly lower than before. This is because the central case assumptions on elasticities involve slightly higher labor supply elasticities for the lower-income classes. Thus, the distortion being relieved is greater than the distortion being aggravated. If the two lowest-income classes are the recipients of the wage subsidy, the MECR is −8.0¢.

The wage subsidy discussed thus far has involved subsidies for all low-income workers. If the subsidy were not made available to household heads in two-earner households, the efficiency should improve, since then the subsidies would be concentrated on secondary earners, who are assumed to have more responsive labor supply. As expected, this plan is more efficient (although only slightly so), with an MECR of −9.2¢.

In Figure 1, I varied the uncompensated labor supply elasticity under the assumption that all groups had the same elasticity. In Table 2, I present some sensitivity analysis with respect to the uncompensated elasticities, under assumptions like those of my central case. The results are similar to those of Figure 1, with a fairly substantial increase in welfare costs as the assumed elasticities rise.

For the cash grant programs, the MECRs depend on both compensated and uncompensated elasticities. For any given uncompensated elasticity, the MECRs increase as the compensated elasticities increase. As the assumed compensated elasticity (and therefore the income effect) rises, this means that cash grants will cause a greater reduction in labor supply. Thus, fewer labor taxes will be collected, and a further increase in tax rates will be necessary in order to balance the government's budget. In the central case, with a total income elasticity of −0.2, the MECR was 80.1¢ for a universal cash grant and 17.7¢ for the notch cash grant program under which the two lowest-income classes receive extra transfers. If we reduce the total income elasticity to −0.1 (holding constant the uncompensated elasticity), these MECRs decrease to 58.4¢ and 14.5¢. If we increase the total income elasticity to −0.3, the MECRs rise to $1.02 and 20.6¢. The wage subsidy results are essentially unaltered by these parameter changes.

Thus far, every simulation reported has been associated with a base case in which 40 percent of GNP are devoted to government spending. If we assume a higher amount of base case spending, we must also assume a higher base case marginal tax rate. Figure 2 shows that the MECRs for cash grants rise as the initial tax rate rises. However, the MECRs for wage subsidies decrease when the inital tax rates rise. We have already seen that the wage subsidy MECR is negative when the distortion being relaxed is greater than the distortion being exacerbated. This effect grows when the inital tax rate on all groups increases. In Figure 2, higher tax rates are brought on by higher exhaustive expenditures. A very similar pattern of results occurs when higher transfer payments are the reason for the higher tax rates.

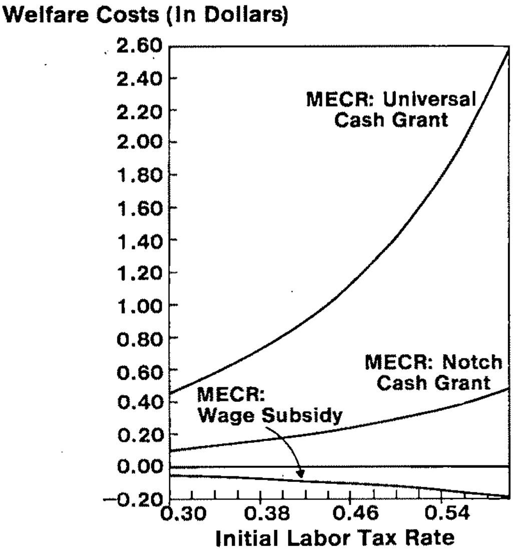


FIGURE 2. MARGINAL EFFICIENCY COST OF REDISTRIBUTION AS A FUNCTION OF INITIAL TAX RATE. CENTRAL CASE LABOR SUPPLY ELASTICITIES ARE USED

Until now, it has been assumed that the initial marginal tax rates facing all groups are the same. As a first approximation, this may not be a bad assumption. In order to generalize the treatment, however, I will now consider several different base case situations. The level of government spending will be the same in each of these as in my "central case." However, instead of labor tax rates of 40 percent on all consumer groups, I consider tax rates on the poorest two classes ranging from 25 percent to 60 percent. The initial tax rate on the higher-income groups must then be adjusted.[6]

The results are shown in Figure 3. When the initial tax rate on the poorest two classes is lower than that on the higher-income groups, the MECR for a wage subsidy is positive (although it is never as great as the MECR for the cash grant programs). When the poor face a tax rate of 25 percent, the MECR for the wage subsidy is 2.3¢. When the poor face higher marginal tax rates than the rich, the MECR is negative, that is, overall efficiency is improved by the transfer. The most dramatic results come when I assume initial tax rates of 60 percent on the two poorest classes. Here, with our central case labor supply elasticities, the MECR is −30.7¢. In other words, we have here a case in which improving the welfare of the "poor" by one dollar costs the "rich" less than 70¢. Throughout this paper, wage subsidies appear as the most efficient means of redistribution. The case is strongest when we believe that the existing marginal tax rates on the poor are high, and/or when we believe that the labor supply elasticities are high.

I have performed some tests of the sensitivity of the model with respect to changes in the distribution of endowments. The MECRs are little affected by changes in the distribution of capital in this model. The distribution of labor endowments does make some difference, however. If labor endowments are assumed to be distributed more equally, the MECRs increase. Consequently, we might

---

[6] Since the bulk of the taxes are paid by the higher-income groups in any event, the adjustments in their tax rates are small. For example, when the poorer groups' tax rate is 25 percent, the rich groups have a rate of approximately 40.652 percent.

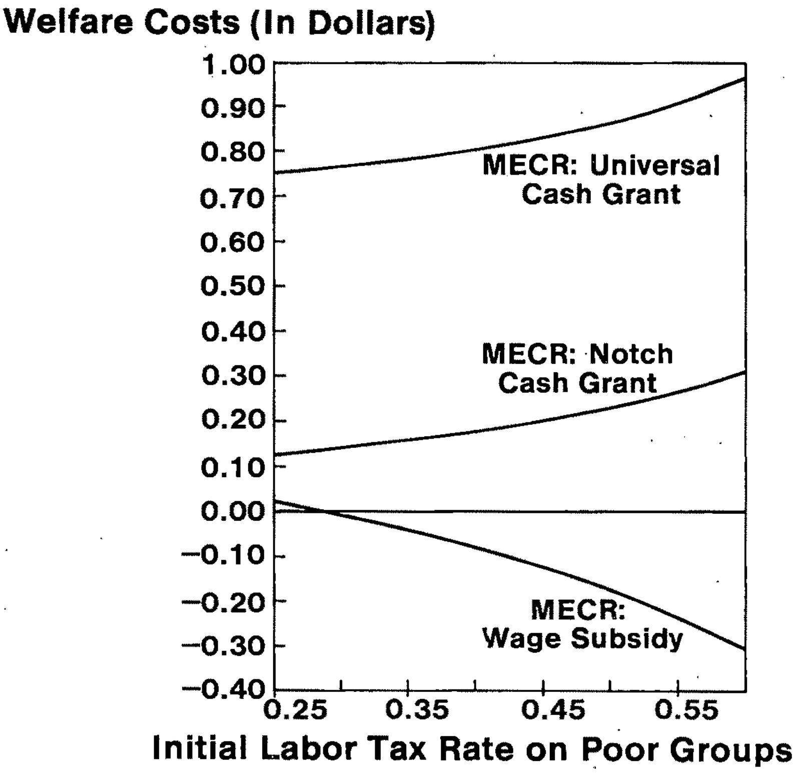


FIGURE 3. MARGINAL EFFICIENCY COST OF REDISTRIBUTION AS A FUNCTION OF INITIAL TAX RATE ON LOW-INCOME CONSUMERS. CENTRAL CASE LABOR SUPPLY ELASTICITIES ARE USED

expect the MECR to be lower for nations that start with a very unequal distribution of income, such as many of the less-developed countries.

We can also consider another way of looking at the marginal efficiency costs. Nicholas Stern (1976) discusses the use of social welfare functions in assessing income distribution. He uses an additive social welfare function that depends upon the utilities of the individual members of society in an isoelastic fashion. One way to write such a function is

$$w = \frac{1}{1-\rho}\sum V^{1-\rho}. \tag{7}$$

If the parameter ρ in equation (7) is zero, then we have a Benthamite social welfare function: we simply add up the individual utilities without assigning any higher weight to lower-income individuals. If ρ is positive, the social welfare function becomes concave, and we place greater weight on the utilities of the poor. In the extreme case of an infinite ρ, we have a Rawlsian (or maximin) social welfare function. For the simulations reported here, we can calculate a critical value of ρ. If society is believed to be characterized by a value of ρ that exceeds the critical value, then the marginal redistribution can be said to have been socially beneficial (even though it clearly is not a Pareto improvement). If society is believed to be characterized by a lower value of ρ, then the extra redistribution would be socially harmful.

I have calculated this critical value of ρ for each of the simulations reported in this paper. In most cases, the critical value is quite small. For instance, in my central case, the critical value is 0.397 for the universal cash grant plan, 0.128 for the notch cash grant plan, and 0.067 for the wage subsidy.[7] For simulations with higher elasticities or higher initial tax rates, the critical values are of course higher. If we assume an initial tax rate of 0.60, the critical value of ρ for the cash grant is 0.927. Note that, for some wage subsidy plans, the critical value of ρ is negative. In these cases, society would have to place greater weight on the utility of the rich than on the utility of the poor in order to not prefer the policy change. I will not speculate here about what the actual ρ is. The construction of an actual social welfare function has well-known problems. The point is simply that one need not be an extreme egalitarian to prefer more redistribution, even when the marginal efficiency costs seem high.

IV. Comparison with Earlier Work

In this section, I report further comparisons between the results presented here and those of Browning and Johnson (1984). They only consider universal cash grant plans, so I cannot compare my calculations with wage subsidies with any results of theirs. Nevertheless, it is interesting to note the disparity between their MECRs for the tax-financed cash grant program and mine. They report (in their Table 8) MECRs ranging from $1.29 to $5.84. They place the greatest emphasis on a combination of parameter values that produce an MECR of $2.49. The simulations in which I place the greatest confidence, on the other hand, yield MECRs of less than one dollar per dollar of additional redistribution.

I ran some additional simulations in an attempt to explain the difference between my results and those of Browning and Johnson. First of all, I adopted their pattern of labor supply elasticities. These elasticities are higher than mine for some groups but lower for others. The weighted average of their uncompensated elasticities is slightly higher than in my central case and the weighted average of their compensated elasticities is slightly lower. Incorporating their elasticities reduces the MECR of a universal cash grant program to 67.1¢. (See the discussion of the effect of changing the total income elasticities, above.) Second, Browning and Johnson assume that the gross wage rate is fixed, that is, they assume an infinitely elastic labor demand curve. To approximate this, I ran my model with all of its standard parameters except for an elasticity of substitution of 20.0. This change (by itself) leads to a slightly higher MECR for the cash grant program of 85.9¢. The wage subsidy MECR remains close to zero. Third, I adopted the pattern of marginal tax rates assumed by Browning and Johnson (1984). These rates are, on average, slightly higher than the 40 percent rate that I use, and they lead to somewhat higher welfare costs. The MECR for a cash grant is 96.4¢ (compared with 80.1¢). The pattern of tax rates adopted by Browning and Johnson is U-shaped. Adopting these tax rates leads to negative marginal efficiency costs (−17.5¢) for the wage subsidy program. A fourth difference stems from the fact that I assume a smaller marginal increase in taxes than do Browning and Johnson. If I increase the marginal change in tax rate from 0.1 percentage points to their 1.0 percentage point, the cash grant MECR rises to 83.4¢.

Next, I put all of these changes in elasticities and tax rates into my model simultaneously. Under this configuration of parameters, the MECR is 89.5¢ for the universal cash grant program, −3.2¢ for the wage subsidy program, and 21.9¢ for the notch cash grant. Thus, these factors do little to explain the difference in the results.

---

[7] In an earlier paper (Ballard, 1987b), the values of ρ for the wage subsidy were usually even a little lower than this. The model used in that earlier paper included only ten consumer groups. Here, once again, the results are complicated by the greater degree of disaggregation employed in this paper, and especially by the fact that some of the population does not work. The wage subsidy does not help them. Thus, even when the MECR is negative, the ρ need not be exceptionally low. This comes out especially in the case where only the poorest class receives the wage subsidy. Nonworkers make up an especially large proportion of this group, and the ρ is 0.405, even though the MECR is only 0.5¢. I also simulated a scheme in which the nonworkers in the lowest-income class receive extra cash grants while the workers receive wage subsidies. In this case, the MECR drops slightly, to −1.9¢, while ρ drops to 0.256.

Another factor explains a portion of the difference between my results and those of Browning and Johnson (1984). In their data set, the upper-income classes have more persons per household, and the lower income classes have fewer persons per household, than in my data set. As suggested above, the MECR of a cash grant program can be expected to rise as the proportion of the marginal benefits going to the high-income groups increases. When I combine the Browning-Johnson allocation of marginal transfers with the other changes described in the previous paragraph, the MECR for the universal tax-financed cash grant rises to \$1.13.[8]

When all the results are put together, I am still unable to reconcile fully the difference between my results and those of Browning and Johnson. Comparisons are difficult, since Browning and Johnson do not use a utility function as such. However, it should be clear that a well-specified computational model may generate MECRs that are substantially below those of Browning and Johnson.

Browning and Johnson (1984) also devote a good deal of attention to comparisons of the changes in money income for the rich and poor groups. They report money income tradeoffs for cash grant plans that are much worse than their MECRs. This is an intuitive result. If we adopt a policy that encourages people to consume more leisure and if they value that leisure, then calculations that concentrate only on money income (and ignore the leisure) are likely to make the policy appear much less efficient. In addition to my MECR calculations, I have calculated this money income tradeoff for all of the policy changes and parameter combinations on which I have reported. I also find that the money income tradeoff is worse than the MECR for cash grant programs. For example, for my central case simulation with marginal tax rates of 40 percent and labor supply elasticities averaging 0.161, the MECR is 80.1¢ but the marginal excess decrease in money income (MEDMI) is \$3.25. However, the large difference between the MECR and MEDMI does not occur when we consider wage subsidy plans. Since the wage subsidy leads to a decrease in leisure for the target population, ignoring the leisure does not make the wage subsidy look worse. For my central case set of parameters, the MECR for a wage subsidy is −8.0¢, and the MEDMI is −17.3¢. In this situation, the poor group's gains in money income actually exceed the losses of the higher-income groups. If the initial marginal tax rate on the lowest two classes is 60 percent, the MECR for a wage subsidy is −30.7¢, and the MEDMI is −36.2¢.

V. Concluding Remarks

In this paper, I have presented the results from a large number of simulations using a simple computational general equilibrium model. The factor proportions, tax rates, and size of government were chosen to be representative of the U.S. economy over the last several years. The standard case elasticity values were selected from the econometric literature, and sensitivity analyses using other values were performed.

For the parameter values considered most believable, the MECR of a universal cash grant financed by increased marginal labor taxes is calculated to be at least 50¢, and may exceed one dollar per dollar of additional redistribution. On the other hand, the marginal efficiency cost of redistributing through a wage subsidy may be negative.

The model within which these results were calculated does have limitations. For example, the model used here abstracts from administrative considerations. The implicit assumption here is that incomes and wages are observed correctly by the administering authority. In the real world, transfer payment programs require the expenditure of resources on administration. These expenditures would probably be the least under a demogrant. The inclusion of administrative costs could increase all of the MECRs, but it would probably reduce the gap between the MECRs for demogrants (which I find to be fairly high) and the MECRs for notch and wage subsidy programs (which I find to be fairly low). A related problem is that, even at this level of disaggregation (which is very great by the standards of CGE models), I distinguish only a part of the heterogeneity in the wage distribution. I capture the presence of lower-earning spouses, but do not account for low-wage teenagers in high-income households. Thus, in the model, the loss of efficiency from subsidizing the wages of low-wage workers in high-income households does not occur. This imparts a further downward bias to the simulated MECRs for wage subsidy schemes.

---

[8] When more of the marginal benefits go to the low-income groups, the MECRs decline. For example, if the marginal transfers are allocated in the same manner as the average transfers in the base case, the MECR falls to 44.3¢ when the other aspects of the Browning-Johnson model are incorporated, and to 37.8¢ when the other parameters of my central case are employed. These lower values emerge because the existing system allocates transfers disproportionately to lower-income households.

Two final qualifications deal with aspects of the model that may lead the MECR calculations for cash grant programs to be biased upward. First, like most CGE models, this model does not allow for unemployment due to macroeconomic rigidities. To the extent that unemployment of this type is suffered disproportionately by low-income individuals, tax-financed cash grant transfers may have significantly lower welfare costs. Second, this model assumes that consumers have no uncertainty about their incomes. In an uncertain world with risk-averse individuals, transfer payments can perform a social insurance role. This could also reduce the MECRs substantially.

The results point toward the need for further consideration of wage subsidies as a policy option. There are many programs with wage subsidy properties, including employment tax credits, the earned income tax credit, and others. Nevertheless, it is reasonable to ask why we do not observe even more wage subsidies. There are several reasons. First, there may be political obstacles. Second, there are a number of practical problems, many of which are discussed in Robert Haveman and John Palmer (1982). Gary Burtless (1985) presents recent evidence on the stigmatizing effect that a wage voucher plan can have. Third, there are some groups in the population who could not be assisted by wage subsidies. Some people are voluntarily at a corner solution in which they do not work in the labor market. For most of this group, a small wage subsidy would not have any effect. Many elderly or disabled persons cannot work under any circumstances. Also, other groups (such as the mothers of small children) may be physically able to work and yet social policy could reasonably be designed so as to encourage them not to work in the labor market. Even if a broad wage subsidy proposal were passed into law, it would still be necessary to have some type of cash grant program as well. A program of wage subsidies for those who can work, combined with cash grants for those who cannot, may be the best program of all. A thorough discussion of one such plan can be found in Kesselman (1973).

It is appropriate to mention some directions for future research along the lines of this simulation study. First, the present study could be extended to incorporate dynamic elements. Second, in a more elaborate model, we could consider different types of tax instruments. Third, one area of very great simplification in the current model is the assumption that all labor types are perfectly substitutable. Future work could allow the target population to have a different skill type from the general population, perhaps using a model with a similar character to that of Martin Feldstein (1973). (This would best be studied in a larger model with heterogeneous output and differences in the consumption patterns of the skill classes.) If cash grant transfer programs cause the recipients to work less, then we would expect that the wages of the occupations dominated by low-income workers would rise. This should reduce the MECRs. The exact extent of the effect would depend on the degree of substitutability of the various labor types, among themselves and with capital.

Computational general equilibrium models are powerful tools for economic policy analysis. They have been used in a variety of applications, but the main emphasis has been on efficiency effects. Less attention has been devoted to distributional issues in general, and transfer payment programs in particular. In this paper, I attempt to reverse this trend by using a CGE model to investigate equity-efficiency tradeoffs. I find that the marginal efficiency costs of redistributing through cash grants are substantial, but my calculations put the efficiency costs well below those found in an earlier study. When increases in the net wages of the poor are the mechanism for redistribution, the efficiency costs are much smaller yet.

REFERENCES

Ballard, Charles L., (1987a) "Marginal Welfare Cost Calculations: Differential Analysis vs. Balanced-Budget Analysis," mimeo., Michigan State University, December 1987.

______, (1987b) "The Marginal Efficiency Cost of Redistribution," Econometrics and Economic Theory Paper No. 8703, Michigan State University, August 1987.

______, (1987c) "The Marginal Efficiency Cost of Redistribution," mimeo., Michigan State University, November 1987.

Berndt, Ernst R., "Reconciling Alternative Estimates of the Elasticity of Substitution," *Review of Economics and Statistics*, February 1976, *58*, 59–68.

Blinder, Alan S., *Toward an Economic Theory of Income Distribution*, Cambridge: MIT Press, 1974.

______ **and Rosen, Harvey S.**, "Notches," *American Economic Review*, September 1985, *75*, 736–47.

Blomquist, N. Sören, "A Comparison of Tax Bases for a Personal Tax," *Scandinavian Journal of Economics*, 1981, *83*(3), 390–407.

______, "The Wage Rate Tax—An Alternative to the Income Tax?," *Scandinavian Journal of Economics*, 1984, *86*(2), 269–85.

Browning, Edgar K., "Alternative Programs for Income Redistribution: The NIT and the NWT," *American Economic Review*, March 1973, *63*, 38–49.

______ **and Johnson, William R.**, "The Trade-Off Between Equality and Efficiency," *Journal of Political Economy*, April 1984, *92*, 175–203.

Burtless, Gary, "Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment," *Industrial and Labor Relations Review*, October 1985, *39*, 105–14.

______, "The Work Response to a Guaranteed Income: A Survey of Experimental Evidence," in Alicia H. Munnell, ed., *Lessons from the Income Maintenance Experiments*, Federal Reserve Bank of Boston Conference Series No. 30, 1987.

Feldstein, Martin S., "On the Optimal Progressivity of the Income Tax," *Journal of Public Economics*, November 1973, *2*, 357–76.

Hamermesh, Daniel S., "The Demand for Labor in the Long Run," in Orley Ashenfelter and Richard Layard, eds., *Handbook of Labor Economics*, Vol. I, Amsterdam: Elsevier, 1986.

Hansson, Ingemar and Stuart, Charles, "Measures of Costs of Taxation: Comment on Browning and Others," mimeo., University of California, Santa Barbara, 1987.

Haveman, Robert H. and Palmer, John L., eds., *Jobs for Disadvantaged Workers: The Economics of Employment Subsidies*, Washington: The Brookings Institution, 1982.

Heckman, James J., Killingsworth, Mark and MaCurdy, Thomas E., "Empirical Evidence on Static Labour Supply Models: A Survey of Recent Developments," in Zmira Hornstein, Joseph Grice, and Alfred Webb, eds., *The Economics of the Labour Market*, London: HMSO, 1981.

Keeley, Michael C., *Labor Supply and Public Policy*, New York: Academic Press, 1981.

Kesselman, Jonathan R., "A Comprehensive Approach to Income Maintenance: SWIFT," *Journal of Public Economics*, February 1973, *2*, 59–88.

______, (1976a) "Egalitarianism of Earnings and Income Taxes," *Journal of Public Economics*, April-May 1976, *5*, 285–301.

______, (1976b) "Tax Effects on Job Search, Training, and Work Effort," *Journal of Public Economics*, October 1976, *6*, 255–72.

______, "Tax Instruments and Economic Equality," *Canadian Journal of Economics*, May 1977, *10*, 289–99.

Killingsworth, Mark R., *Labor Supply*, New York: Cambridge University Press, 1983.

Kimbell, Larry J. and Harrison, Glenn W., "General Equilibrium Analysis of Regional Fiscal Incidence," in Herbert E. Scarf and John B. Shoven, eds., *Applied General Equilibrium Analysis*, New York: Cambridge University Press, 1984, 275–313.

Stern, Nicholas H., "On the Specification of Models of Optimum Income Taxation," *Journal of Public Economics*, July-August 1976, *6*, 123–62.

U.S. Department of Commerce, Bureau of Economic Analysis, "Revised Estimates of the National Income and Product Accounts," *Survey of Current Business*, July 1982, *62*, No. 7.

______ **Bureau of the Census,** *Current Population Reports*, Series P-60, No. 129, "Money Income of Families and Persons in the United States: 1979," Washington: USGPO, 1981.

______, *Current Population Reports*, Series P-23, No. 126, "Estimating After-Tax Money Income Distribution Using Data from the March Current Population Survey," Washington: USGPO, 1983.

______, *1980 Census of Population*, Vol. I, *Characteristics of the Population*, Chapter D: Detailed Population Characteristics, Part I: United States Summary, PC80-1-D1-A, Section A: United States, March 1984. Washington: USGPO, 1984.

Wagstaff, Peter, "A Benthamite Wages Policy," *Review of Economic Studies*, October 1975, *42*, 571–80.

Wildasin, David E., "On Public Good Provision with Distortionary Taxation," *Economic Inquiry*, April 1984, *22*, 227–43.

Zeckhauser, Richard J., "Optimal Mechanisms for Income Transfers," *American Economic Review*, June 1971, *61*, 324–34.

Altruism Within the Family Reconsidered: Do Nice Guys Finish Last?

By B. Douglas Bernheim and Oded Stark*

We criticize the view that altruism either increases the benefits of group interactions or improves the allocation of resources within families. Altruism can alter the social utility possibility frontier in surprising and sometimes unfortunate ways. Altruism also often entails exploitability and therefore causes family members to behave in ways that leave all parties worse off. In addition, altruists have difficulty enforcing agreements since they may be extremely reluctant to punish betrayals.

Recent work on the economics of family behavior has emphasized the relationships between altruism, voluntary transfers, and internal resource allocation.[1] Two themes have emerged from this literature. First, since altruists internalize externalities, families that contain altruists should allocate resources efficiently. This principle is most clearly enunciated in Gary Becker's "Rotten Kid Theorem" (see, for example, Becker, 1981).[2] Second, altruism increases the benefits of group interaction. One consequence of this principle is that one ought to discover altruistic linkages between spouses.[3]

It is now well-established that the first of these results is rather special. Allocative inefficiencies may persist despite the existence of altruistically motivated resource transfers (see Jack Hirshleifer, 1977; B. Douglas Bernheim, Andrei Shleifer, and Lawrence Summers, 1985; or Theodore Bergstrom, 1986). This is especially true in intertemporal contexts, where altruism may give rise to phenomena such as the "Samaritan's dilemma" (Neil Bruce and Michael Waldman, 1986; Bergstrom, 1986).[4]

In this paper, we take the critique of altruism as a constructive social force a step further. In particular, we identify a variety of circumstances in which members of a group would actually prefer to interact with less altruistic individuals, and in which the efficiency of resource allocation is inversely related to the prevailing degree of altruism. This conclusion is clearly much stronger than previous results, which simply indicated that the presence of altruism need not eliminate all social inefficiencies.

In the following sections, we discuss a variety of reasons for suspecting that altruism might be a counterproductive social force. Our first point (raised in Section I) is non-behavioral: altruism can alter the social utility possibility frontier in surprising and sometimes unfortunate ways. The intuition is simple: if A loves B and B is unhappy, then A's love may cause A to be unhappy as well. Worse yet, if B also loves A, then A's love

---

*Department of Economics, Stanford University, Stanford, CA 94305; and Harvard University, Cambridge, MA 02138 and Bar-Ilan University, Ramat-Gan, 52100, Israel, respectively. We would like to thank James Andreoni, Theodore Bergstrom, Robert Pollak, Robert Willis, and several anonymous referees for helpful comments. The first author would also like to thank the National Science Foundation for financial support (National Science Foundation grant no. SES-8607630).

[1] See especially Gary Becker, 1974a, 1981.

[2] The Rotten Kid Theorem says the following: suppose that selfish family members are linked through resource transfers to an altruistic head-of-household. Then each member will act as though his objective is to maximize the total well-being of the family.

[3] See Becker, 1973, 1974b, 1981; or Becker, Elisabeth Landes, and Robert Michael, 1977.

[4] The Samaritan's dilemma refers to the fact that a selfish individual may impoverish himself through extravagance, knowing that an altruist will provide for him in the future. Our analysis in Section II is based upon a model in which this possibility arises.

for B would then make B all the more unhappy.

Our next two points concern behavior. We argue that in a variety of situations, altruism entails exploitability, and therefore causes family members to behave in ways that leave all parties worse off (this observation is no doubt well understood by most parents). In Section II, Part A, we consider an environment in which altruism gives rise to incentive problems (specifically, the Samaritar's dilemma). For a robust set of cases, we show that altruists may take inefficient actions in order to preempt exploitation; in cases where altruism is stronger, these actions must be more extreme, and so the resulting inefficiency is greater. In Section II, Part B, we consider the possibility that family members may succeed in establishing efficient resource allocation by punishing selfish behavior. We show that a rise in the level of altruism may lead to the exclusion of certain punishment strategies on the grounds of credibility. Since an altruist may be extremely reluctant to punish betrayals, others may perceive him as a "softy." This may severely limit his ability to participate in cooperative agreements that are beneficial to all members of the group.

While these findings are generally negative in nature, we do obtain one very general positive result. Specifically, in Section II, Part C, we consider an extremely broad class of environments, and demonstrate that sufficiently high levels of altruism almost always (in a sense made precise) lead to efficient resource allocation. However, since the relationship between altruism and economic performance need not be monotonic (as shown in Section II, Parts A and B), this observation does not validate the view that altruism is generally a constructive social force.

Although we emphasize the connections between this work and the literature on family behavior, our analysis also has much broader implications concerning the role of altruism in society, and the evolution of social conscience. In this respect, it is related to work by Robert Axelrod (1984) and Robert Frank (1986). We briefly discuss some of these implications in the concluding section.

I. Altruism and the Utility Possibility Frontier

Throughout this paper we focus our attention on situations involving two distinct types of individuals, labeled A and B. Although we assume that these individuals interact in pairs, the principles discussed here easily generalize to larger groupings. For concreteness, it is helpful to cast our discussion in terms of the selection of spouses. Accordingly, we think of type As as "males," and Bs as "females." Our strategy in this section is to analyze the preferences of an arbitrarily selected type B individual over the set of potential partners. We index type As by the parameter θ, and take θ to lie on the unit interval. Below we describe the relationship between θ and personal characteristics.

We assume that the utility of type B is given by some function $U(C,c,\theta)$, where C is B's own consumption vector, and c is that of her partner. Supposing that she pairs with θ, his utility will be given by $u_\theta(C,c)$. Our central assumption is that it is possible to represent these functions as follows:

$$U(C,c,\theta)=\Gamma V(C)+Bu_\theta(C,c) \tag{1}$$

$$u_\theta(C,c)=\gamma(\theta)v(c)+\beta(\theta)U(C,c,\theta), \tag{2}$$

where $V(\cdot)$ and $v(\cdot)$ are increasing functions, and where Γ, B, $\gamma(\theta)$, and $\beta(\theta)\geqq 0$ (i.e., we rule out masochism and envy).[5] Furthermore, we assume that this representation is unique.

This formulation of altruism is standard (see, for example, Robert Barro, 1974 and Becker, 1974a). Furthermore, equations (1) and (2) have very nice interpretations. Utility arises from two sources: the felicity associated with an individual's own consumption, and the utility of his or her partner. The relative weight attached to an individual's partner reflects the intensity of altruism.

---

[5] These restrictions do not necessarily follow from natural assumptions about $U(\cdot)$ and $u_\theta(\cdot)$. It is nevertheless quite common to specify preferences through a system of equations like (1) and (2), and to simply assume that the restrictions are satisfied. This issue is of no real importance in the current context. Indeed, since we merely wish to establish the possibility that outcomes may be perverse, it is more compelling to confine our attention to as standard and well-behaved a set of environments as possible.

Note that we have taken Γ, B, $v(\cdot)$, and $V(\cdot)$ to be independent of θ. Thus, the type B individual cares equally about all potential partners, and type As differ only in that some care more for the type B than others. This assumption is appropriate since our object is to analyze an individual's preferences over potential partners who are the same in all respects, except that some are more altruistic than others. We can arbitrarily assign the index θ such that larger values correspond to greater altruism. Thus, without loss of generality we take $\beta'(\theta) > 0$. We also assume $\gamma'(\theta) \leqq 0$—altruistic individuals do not attach more importance to their own consumption than do more selfish individuals.

Thus far, we have said nothing about a couple's feasible consumption set. Implicitly, we take this to be the same for all potential couples. For each feasible consumption profile, we analyze the relationship between utility and the choice of a partner. This allows us to draw conclusions about the effect of altruism on the utility possibility frontier. Below we study in detail the impact of altruism on choices within the feasible set.

Having fixed arbitrary values of C and c, we solve (1) and (2) to express type B's utility directly in terms of felicities and her choice of a partner:

$$(3)\quad U(C,c,\theta) = [\Gamma V(C) + B\gamma(\theta)v(c)](1-\beta(\theta)B)^{-1}.$$

If we take the primitive utility function $U(\cdot)$ to be increasing in c and C, then under the rest of our assumptions it must be the case that $\beta(\theta)B < 1$.

To determine the relationship between type B's utility and her choice of a partner, we differentiate equation (3) with respect to θ:

$$(4)\quad \frac{\partial U}{\partial \theta} = \{v(c)[\gamma'(\theta)[1-\beta(\theta)B] + B\gamma(\theta)\beta'(\theta)] + V(C)\Gamma\beta'(\theta)\} \cdot B(1-\beta(\theta)B)^{-2}.$$

The derivative has the same sign as the expression in curly brackets. Thus, $\partial U/\partial\theta > 0$ (type B prefers a more altruistic partner) if and only if

$$(5)\quad V(C) > v(c)\phi(\theta),$$

where

$$(6)\quad \phi(\theta) = \{-\gamma'(\theta)[1-\beta(\theta)B] - B\gamma(\theta)\beta'(\theta)\}[\Gamma\beta'(\theta)]^{-1}.$$

Note in particular that $\phi(\cdot)$ does not depend upon the felicity levels $V(C)$ and $v(c)$. Thus, for any consumption bundle we can determine whether or not type B would prefer a more altruistic partner simply by inspecting the ratio of felicities.

Two features of equations (5) and (6) deserve emphasis. First, if type B's felicity is sufficiently low, she will prefer a less altruistic partner. In this case, nice guys do indeed finish last. The explanation is quite simple. An altruistic type A would be depressed by his partner's low level of felicity. Since the type B cares about her partner, she would in turn be disturbed by the fact that she has made him unhappy.

Second, note that $\phi(\theta)$ may be either positive or negative (the two terms in curly brackets have opposite signs). If it is positive, then when type As are sufficiently happy ($v(c)$ is large) the type B will again prefer a less altruistic partner. Intuitively, since she cares about her partner, she would rather he place more importance on his own high level of felicity than on her comparatively low level of utility. To take an example, suppose that $\gamma(\theta) = 1-\beta(\theta)$ and $\Gamma = 1-B$, so that utility is a simple weighted average of own felicity and spouse's utility. Then it is easy to check that $\phi(\theta)=1$. Whether or not the type B prefers a more altruistic partner will then depend simply upon whether $V(C)$ is larger or smaller than $v(c)$.

If $\phi(\theta)$ is negative, then the result described in the preceding paragraph is reversed: when $v(c)$ is sufficiently small (type As are unhappy), the type B will prefer a less altruistic partner. Note that this cannot occur unless either $v(c)$, $V(C)$, or both are negative. To take an example, suppose that $\gamma(\theta)=\Gamma=1$, so that altruism does not affect the absolute importance attached to own consumption. Then $\phi(\theta)=-B$. When $V(C) < -v(c)B$, the type B's utility is necessarily negative. It is therefore quite natural that she would not want her partner to attach greater importance to her well-being.

One final point deserves mention. Consider two type As, θ and θ'. From (1), it follows that $U(C,c,\theta) > U(C,c,\theta')$, if and only if $u_\theta(C,c) > u_{\theta'}(C,c)$. Thus, the type B's optimal choice leads to a partnership that is happier all around. In particular, if she prefers a selfish partner, then selfish type As will be happier as her partner. To state this in a slightly different way, suppose that all type As have the same level of reservation utility—unless they achieve this level, they will not form a partnership. Then if a type B prefers a selfish partner, only sufficiently selfish type As will be willing to pair with her.

The reader might well object to this analysis on the grounds that we have taken behavior as fixed. In practice, the type B might always prefer a more altruistic partner, since he would be inclined to enhance her felicity through transfers and other activities. We agree that it is essential to incorporate the impact of altruism on behavior. However, it is easy to see that in a robust set of cases, the effects discussed here will dominate behavioral responses.

Consider the following example. Let $\gamma(\theta)=1-\beta(\theta)$ and $\Gamma=1-B$. Suppose that each type A is endowed with wealth W, and that the type B has no resources of her own. In a partnership, the type A decides how to divide consumption between himself and his partner (i.e., he chooses C and c subject to $C+c\leq W$). Let $V(C)=\ln(C\mu)$ with $\mu>0$, and $v(c)=\ln(c)$. It is easy to verify that individual θ would consume $W\alpha(\theta)$ himself, where $\alpha(\theta)=[1-\beta(\theta)][1-\beta(\theta)B]^{-1}$, transferring the residual to his partner. We can then write type B's utility indirectly as a function of θ (denoted $\hat{U}(\theta)$), where it is understood that we have substituted for type A's optimal choice. Differentiation yields

$$\frac{d\hat{U}}{d\theta}=\alpha'(\theta)\left\{B\ln\left[\frac{\alpha(m)}{(1-\alpha(m))\mu}\right]-\left[\frac{1-B}{1-\alpha(m)}\right]\right\}.$$

We are primarily concerned with the case where $B<1$ (otherwise we would have $\Gamma<0$, and θ's optimal choice would be to consume everything himself). It follows that $\alpha'(\theta)<0$. Thus, when the first term in curly brackets is sufficiently small, the type B individual prefers a more altruistic partner. However, μ can be chosen arbitrarily without affecting behavior. By taking μ to be sufficiently small, we can make this term arbitrarily large, thereby producing the result that the type B prefers a less altruistic partner, despite the fact that he will provide her with fewer resources.

II. Altruism and Behavior

In the simple example described at the end of the preceding section, altruism had a beneficial impact on transfers (unfortunately, this effect was simply dominated for small values of μ). We now describe several kinds of environments in which the presence of actual or potential resource transfers causes altruism to have counterproductive behavioral effects. In contrast to the approach adopted in Section I, we take partnerships as given, and compare the behavior of couples who differ only with respect to the strength of their altruistic motives.

Accordingly, in this section it is convenient to assume that there are only two individuals and to identify each individual i with his type, A and B. We assume that each agent has preferences of the form given by equations (1) and (2), and solve for utilities in terms of felicities, as in equation (3). Since we take partnerships as given and focus on behavior, we are only concerned with ordinal preferences over consumption. Without loss of generality, we may therefore normalize utility so that agent i places a weight of $1-\alpha_i$ on his own felicity and a weight of α_i on that of his partner:

$$(7)\quad u_i(C_i, C_j) = (1-\alpha_i)V_i(C_i) + \alpha_i V_j(C_j),$$

where u_i is the utility, V_i the felicity, and C_i the consumption of $i = A, B$, and where $j \neq i$. Throughout, we assume that $\alpha_A + \alpha_B \leqq 1$, thereby ruling out the complications that arise in cases of "excessive" altruism. Clearly, $\alpha_A + \alpha_B = 1$ corresponds to the case of "perfect" altruism, in which the agents' interests are completely harmonious.

In Parts A and B we describe some specific environments in which consumption is determined by strategic play. We demonstrate that in these environments selfish couples may well succeed in allocating resources more efficiently, thereby achieving strictly higher levels of felicity than altruistic couples. In Part C, we isolate some general conditions under which altruistic couples fare better.

A. *Incentive Problems*

We consider a version of the Samaritan's dilemma in which the two agents, A and B, live for two periods. Each agent i is endowed at the beginning of period 1 with resources w_i. A consumes c_A^t in each period t; we use c_B^t to denote B's consumption. A and B make decisions in the following order: first, A selects c_A^1, then B selects c_B^1, then A selects c_A^2, transferring his residual resources $(w_A - c_A^1 - c_A^2)$ to B, and finally B consumes her residual resources $(w_B - c_B^1 + w_A - c_A^1 - c_A^2)$.[6] For analytic simplicity, we specify felicity to be loglinear, and assume that the interest rate and pure rate-of-time preference are both zero. Finally, we assume that A cares about B—a fact which B is aware of—but not *vice versa*.[7] Accordingly, we adopt the following parameterization of equation (7):

$$u_A = (1-\alpha_A)\sum_{t=1,2}\ln c_A^t + \alpha_A \sum_{t=1,2}\ln c_B^t,$$

and

$$u_B = \sum_{t=1,2}\ln c_B^t.$$

This environment determines a game of perfect information. We solve for the subgame-perfect (closed loop) Nash equilibrium (see Reinhard Selten, 1975) through backward recursion.

Clearly, in the final stage, B will simply consume all remaining resources. Accordingly, given c_A^1 and c_B^1, A will want to allocate the fraction α_A of total second-period resources to B, and keep the remainder for himself. He may, of course, be constrained in this regard, in that he cannot appropriate B's wealth. Thus,

$$(8)\quad c_A^2 = \min\{(1-\alpha_A)(w_A + w_B - c_A^1 - c_B^1),\ w_A - c_A^1\},$$

$$(9)\quad c_B^2 = \max\{\alpha_A(w_A + w_B - c_A^1 - c_B^1),\ w_B - c_B^1\}.$$

---

[6] Although this sequential order of play is quite special, the result described in this section is very general. If, for example, A and B choose simultaneously in each period, then one could produce the same result in a three-period model. Note also that we have restricted A from making transfers in period 1 ("gifts," as opposed to period 2 "bequests"). Implicitly, we assume that agents prefer to transfer resources later in life, perhaps because they wish to maintain reserves against adverse contingencies, because of taxes, or because of other incentive problems (see, for example, Bruce and Waldman, 1986). It is clearly possible to obtain the same result in a model where gifts are treated explicitly. We opt for the simpler sequential model without period 1 transfers in order to illustrate the exploitability of an altruist in the most transparent setting.

[7] This is for analytic convenience only. Two-sided altruism does not alter the basic principles.

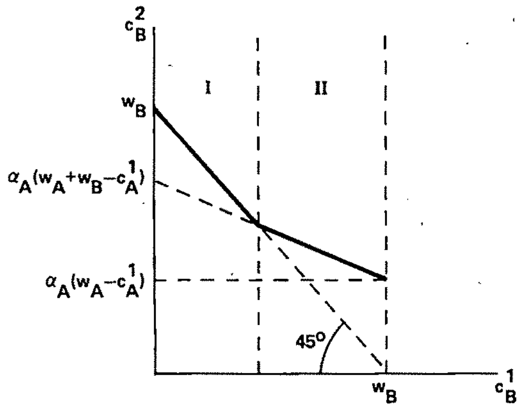


FIGURE 1. *B*'S EFFECTIVE BUDGET CONSTRAINT

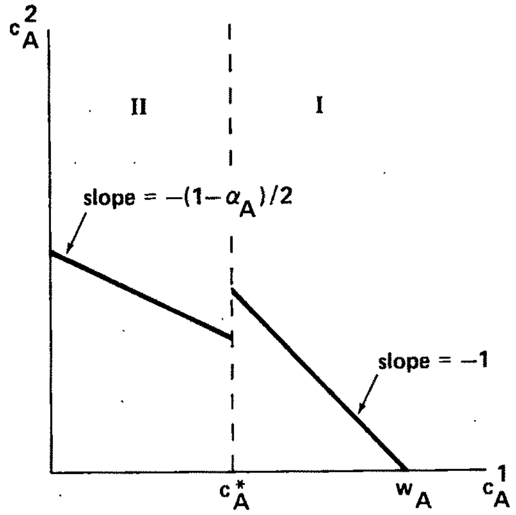


FIGURE 2. *A*'S FEASIBLE CONSUMPTION CHOICES

We now move back an additional step, and consider *B*'s choice of c_B^1. Equation (9) determines *B*'s effective budget constraint, which we illustrate in Figure 1. Note that we have identified two regimes, which we designate I and II. If *B* chooses to consume on part I of her budget constraint, then

$$c_B^{1,\mathrm{I}} = w_B/2. \tag{10}$$

If, on the other hand, she chooses to consume on part II, then

$$c_B^{1,\mathrm{II}} = \left(w_A + w_B - c_A^1\right)/2. \tag{11}$$

It is essential to determine whether *B* chooses $c_B^{1,\mathrm{I}}$ or $c_B^{1,\mathrm{II}}$. Accordingly, we identify the level of c_A^1 for which she is just indifferent by solving the equation

$$2\ln(w_B/2) = \ln\left[\left(w_A + w_B - c_A^1\right)/2\right] + \ln\left[\alpha_A\left(w_A + w_B - c_A^1\right)/2\right].$$

The solution, c_A^*, is given by

$$c_A^* = w_A + w_B\left(1 - \alpha_A^{-1/2}\right). \tag{12}$$

When $c_A^1 > c_A^*$, *B* will select $c_B^{1,\mathrm{I}}$. When $c_A^1 < c_A^*$, *B* will choose $c_B^{1,\mathrm{II}}$.

We now consider *A*'s initial decision concerning c_A^1. From equations (8) through (11), we infer that

$$c_A^2 = \begin{cases} w_A - c_A^1 & \text{for } c_A^1 \geqq c_A^* \\ (1-\alpha_A)\left(w_A + w_B - c_A^1\right)/2 & \text{for } c_A^1 < c_A^* \end{cases} \tag{13}$$

$$c_B^1 = \begin{cases} w_B/2 & \text{for } c_A^1 \geqq c_A^* \\ \left(w_A + w_B - c_A^1\right)/2 & \text{for } c_A^1 < c_A^* \end{cases} \tag{14}$$

$$c_B^2 = \begin{cases} w_B/2 & \text{for } c_A^1 \geqq c_A^* \\ \alpha_A\left(w_A + w_B - c_A^1\right)/2 & \text{for } c_A^1 < c_A^*. \end{cases} \tag{15}$$

Equation (13) merits further attention, in that it describes *A*'s feasible consumption locus. Note that

$$\begin{aligned} w_A - c_A^* &= -\,w_B\left(1 - \alpha_A^{-1/2}\right) \\ &\geqq -\,w_B\left(\alpha_A^{1/2} - \alpha_A^{-1/2}\right)/2 \\ &= (1-\alpha_A)\left(w_A + w_B - c_A^*\right) \end{aligned}$$

(with strict inequality for $\alpha_A < 1$), so that the set of *A*'s potential consumption choices is discontinuous at c_A^*, as shown in Figure 2. As before, we identify regimes I and II.

From inspection of Figure 2, it is evident that there are three types of equilibria. First, *A*'s choice might be interior to regime I. This occurs when A's altruism is so weak that A wishes to extract some of B's endowment from B. The presence of altruism does not alter incentives through transfers, so equilibria in regime I are Pareto efficient. Second, A's choice might be interior to regime II. This occurs when A's altruism is sufficiently strong to motivate transfers. The Samaritan's dilemma results: since B faces an implicit tax on saving, his first-period consumption is inefficiently high. Third, A's choice might be at the point of discontinuity. It is this possibility that we wish to investigate further.

Clearly, one cannot have an equilibrium at the point of discontinuity unless $c_A^* \geqq w_A/2$ (otherwise A would prefer to divide his resources evenly between the two periods). Using our formula for c_A^* (equation (12)), this implies

$$\alpha_A \geqq (1 + w_A/2w_B)^{-2} \equiv \underline{\alpha}_A.$$

Suppose this inequality holds. Would A prefer to choose c_A^*, or some consumption level within regime II? While we have been unable to obtain a general answer to this question, it is clear that there is a robust set of cases for which A prefers c_A^*. For example, if $\alpha_A = \underline{\alpha}_A$ and $w_A = w_B = 1$, then A can obtain a utility level of -1.39 at the point of discontinuity. On the other hand, by maximizing his utility subject to the relevant portions of equations (13) through (15), one can verify that he can do no better than -1.48 in regime II. Since the utility associated with both options responds continuously to changes in α_A, we conclude that there is some $\bar{\alpha}_A > \underline{\alpha}_A$, such that if $\alpha_A \in [\underline{\alpha}_A, \bar{\alpha}_A]$, the equilibrium lies at the point of discontinuity.

Suppose then that $\alpha_A \in (\underline{\alpha}_A, \bar{\alpha}_A)$, and consider the effect of raising α_A. More precisely, compare the behavior of two different couples where the type A individuals have distinct values of α_A within this interval. We know that for small changes, the equilibrium remains at the point of discontinuity. Thus,

$$\frac{dc_A^1}{d\alpha_A} = -\frac{dc_A^2}{d\alpha_A} = \frac{dc_A^*}{d\alpha_A} = \frac{w_B\alpha^{-3/2}}{2} > 0,$$

and

$$\frac{dc_B^1}{d\alpha_A} = \frac{dc_B^2}{d\alpha_A} = 0.$$

How does this small change affect the efficiency of the resulting allocation? To answer this question, it is appropriate to focus on felicities. Since $v_A(c_A^1, c_A^2) = \ln c_A^1 + \ln c_A^2$, we have

$$\frac{dv_A}{d\alpha_A} = \left[\frac{1}{c_A^*} - \frac{1}{w_A - c_A^*}\right]\frac{dc_A^*}{d\alpha_A} < 0$$

(since $c_A^* > w_A/2$). Thus, couples with higher levels of altruism (α_A) will allocate resources less efficiently. Within this regime, they achieve a lower level of felicity for the type A individual, and the same level for the type B individual.

It is worthwhile spending a moment to understand this result. The intuition is as follows. A knows that B may consume a large share of her resources in period 1 with the intent of freeloading in period 2. To prevent this, A must find some way of committing himself not to support B in period 2. In our model, A can accomplish this only by consuming a large fraction of his own resources in period 1. Freeloading then becomes less lucrative for B, who instead chooses to support herself. In a couple where A is more altruistic toward B, A must take more extreme measures in order to make freeloading unattractive to B. That is, he must consume an even larger share of his resources in period 1. Since the first-period consumption of type A individuals is too large to begin with, this causes a further loss of efficiency.

B. *Cooperation*

In this part, we study the ability of social groups to achieve cooperative outcomes by punishing selfish behavior. Since family relationships are usually ongoing, it is natural to assume that family members interact in a sequence of situations. Failure to cooperate at any point in time may lead to subsequent retaliation. However, it is important to require that the threat of retaliation be credible, in the sense that the retaliating party would actually find it in his interest to carry out the prescribed punishment.

We consider a simple stylized model of an ongoing relationship between two individuals. Each agent, A and B, wishes to consume two goods, x and y, so that $c_i = (x_i, y_i)$. Both goods are nonstorable. In each period, $t = 1, 2, \ldots, \infty$, A is endowed with $k - \mu$ units x and μ units of y, while B is endowed with $k - \mu$ units of y and μ units of x. We also assume that A and B cannot exchange x for y (or vice versa) on open markets: at least one of these goods represents a resource or service (for example, loving care) that is provided exclusively through families. Throughout, we will take $\mu > 0$ to be very small, so that each agent virtually monopolizes one good.[8] The preferences of each agent i are given by equation (7). We specify felicity as a function of own consumption

$$v_i(x_i, y_i) = (x_i y_i)^{\gamma},$$

where $0 < \gamma < 1/2$. The restriction that $\gamma < 1/2$ assures that u_i is quasi-concave in the consumption of individuals A and B. Henceforth, we will also impose the restriction that $\alpha_A = \alpha_B = \alpha$ (i.e., symmetric altruism).

In analyses of repeated games, it is standard practice to enforce cooperative outcomes through "Nash reversion" (see, for example, James Friedman, 1977). That is, agents play cooperative actions on the equilibrium path. However, if either player ever deviates from his assigned choice, then the equilibrium prescribes the static Nash solution in every subsequent period. Such punishments are credible in that each agent willingly participates in the punishment, given the other's participation (i.e., the equilibrium is subgame perfect). While more severe punishments are frequently available (see Dilip Abreu, 1986), Nash reversion is by far the simplest and most widely discussed method of enforcing cooperation. We employ it here.

It is therefore necessary to begin by characterizing static Nash outcomes. Since μ is very small, agent A will, in effect, determine the division of x and B will determine the division of y. Suppose that agent B chooses to consume $y_B \equiv \lambda_B k$ (i.e., λ_B is the fraction of the aggregate endowment of y that B retains for himself). Then A solves

$$\max_{\lambda_A} (1-\alpha)\left[\lambda_A(1-\lambda_B)\right]^{\gamma} k^{2\gamma} + \alpha\left[(1-\lambda_A)\lambda_B\right]^{\gamma} k^{2\gamma},$$

subject to $0 \leqq \lambda_A \leqq 1 - \mu/k$ (where λ_A is the fraction of his endowment that A retains for himself). Differentiation reveals that

$$(16) \qquad (1-\alpha)\left(\frac{1-\lambda_B}{\lambda_B}\right)^{\gamma} = \alpha\left(\frac{1-\lambda_B}{\lambda_B}\right)^{\gamma-1}.$$

Since the first-order condition for B is completely symmetric, it is natural to look for symmetric equilibria. Setting $\lambda_A = \lambda_B$ (and accounting for boundary conditions), we find that the equilibrium choice, λ^N, is given by

$$(17) \qquad \lambda^N = \min\{1-\alpha, 1-\mu/k\}.$$

Note that as α approaches 1/2 (perfect altruism), players split all resources equally (the symmetric cooperative outcome). It is straightforward to show that (λ^N, λ^N) is the unique Nash equilibrium given the parameter restrictions described above.[9] Utility levels at the static Nash solution are given by

---

[8] When $\mu = 0$, there are always three static Nash equilibria, one of which is symmetric (as in the text). The other two equilibria have $x_i = y_i = 0$ for some i. This is a rather peculiar outcome—i gives everything away simply because retaining it has no effect on his felicity. When $\mu > 0$, i always has an incentive to retain some fraction of his endowment.

[9] One can see this by taking logs of equation (16) and its counterpart for B. This yields a nondegenerate system of two linear equations in two unknowns. Thus, there can be only one interior solution. One can check directly that boundary cases cannot be equilibria.

$$(18)\qquad u^N = \left[\lambda^N(1-\lambda^N)\right]^{\gamma} k^{2\gamma}.$$

Our objective is to determine the feasibility of achieving cooperative equilibrium outcomes. For simplicity, we focus on the case of symmetric distribution: $x_A = x_B = y_A = y_B = k/2$ (note that this is Pareto-efficient irrespective of α). Associated utility is given by

$$(19)\qquad u^C = \left(\frac{1}{4}\right)^{\gamma} k^{2\gamma}$$

(the superscript C denotes cooperation).

Suppose agent i contemplates deviating from the cooperative allocation. Using equation (16), one can show that it is optimal for him to select

$$(20)\quad \lambda_i^D = \min\left\{\left[1+\left(\frac{\alpha}{1-\alpha}\right)^{1/(1-\gamma)}\right]^{-1},\; 1-\mu/k\right\}$$

By deviating in this way, i can obtain a current payoff of

$$(21)\qquad u^D = (1-\alpha)\left[\lambda^D/2\right]^{\gamma} k^{2\gamma} + \alpha\left[(1-\lambda^D)/2\right]^{\gamma} k^{2\gamma}$$

(the superscript D denotes deviation).

When is cooperation sustainable through Nash reversion? Agent i must believe that the short-run gain from deviating does not exceed the discounted value of future losses. That is,

$$(22)\quad [u^D - u^C] \leqq \frac{\delta}{1-\delta}[u^C - u^N],$$

where δ is the rate at which agents discount future utility.[10] It follows that cooperation is sustainable as long as

$$(23)\qquad \delta \geqq [u^D - u^C]/[u^D - u^N].$$

Intuitively, cooperation requires that each agent place sufficient weight on the future, so that his fears of retaliation dominate any inclination toward short-run opportunism.

How does altruism affect the prospects for cooperation? It helps to build intuition by considering the case of α close to zero (pure selfishness) and μ very small. From equation (17), it is clear that λ^N is close to 1. Thus, A's (B's) marginal utility from y (x) approaches infinity. Since a change in α induces a first-order change in λ^N, the effect on u^N will be extremely large. Accordingly, the right-hand side of (22) falls very sharply with α (in the limit, the derivative is infinite). Since altruism improves the static noncooperative outcome, it reduces the severity of credible punishments. An increase in altruism will also make deviations less desirable. However, from equations (20) and (21), it is possible to show that this effect is of the same order of magnitude as the change in α. Thus, the right-hand side of (22) falls by more than the left, and cooperation becomes *more* difficult to sustain.

To illustrate this point more concretely, we consider a numerical example. Let $\mu/k = 0.01$ and $\gamma = 0.25$. For values of $\alpha \in [0, 1/2)$ (recall that with $\alpha_A = \alpha_B$, $\alpha = 1/2$ represents perfect altruism), we calculate δ_α^*, the minimum value of δ that satisfies inequality (23). δ_α^* is the *lowest* discount factor for which one can sustain cooperation. Thus, lower values of δ_α^* imply that it is easier to sustain cooperation. Using equations (17) through (21), we obtain the entries in Table 1. Notice that δ_α^* rises *monotonically* with α. Regardless of where one starts, an increase in the level of altruism makes cooperation more difficult. Suppose, for example, that $\delta = 0.35$. For a couple with $\alpha = 0.10$, it is possible to sustain complete cooperation. However, a more altruistic couple with $\alpha = 0.15$ cannot achieve the cooperative outcome.

It is important to emphasize that one can also construct examples in which altruism has the opposite effect. Yet this discussion

---

[10] That is, total utility is given by $\sum_{t=0}^{\infty}\delta^t u_{it}$.

TABLE 1—THRESHOLD DISCOUNT FACTORS

Index of Altruism α	Threshold Discount Factor δ_α^*
0	.251
.05	.301
.10	.334
.15	.354
.20	.369
.25	.379
.30	.387
.35	.393
.40	.397
.45	.399

serves to emphasize a general point: when altruism improves static noncooperative outcomes, it lessens the severity of credible punishments. An altruist may well be perceived as a "softy" and his threats may not be taken seriously. Most parents should be familiar with this problem. In addition, such considerations may well explain why some alliances (for example, the Warsaw Pact) are perceived to be stronger than others (for example, NATO): a stronger commitment to the goals of an alliance (deeper caring for the security of the allies) may actually make the alliance less effective.[11] While altruism may also dampen incentives to deviate, the first effect dominates in a robust class of environments.

C. *A Positive Result*

In view of the examples presented in Parts A and B, it is natural to wonder whether one can make any general statements about the effects of altruism on economic performance. Surely, since perfect altruism ($\alpha_A + \alpha_B = 1$) induces perfect harmony (in the sense that strategic problems disappear and cooperation prevails), sufficiently high levels of altruism ($\alpha_A + \alpha_B$ close to 1) should also improve equilibrium outcomes. Unfortunately, it is relatively easy to construct counterexamples in which even this relatively weak conjecture is false.[12] Yet these examples seem quite contrived, in that they tend to rely on coincidental features of payoff functions. We now formalize this reasoning.

Consider the set of two-agent environments (agents A and B) in which consumption is determined as a function of strategic choices, x_A and x_B. The choice sets X_A (for A) and X_B (for B) have cardinality I and J, respectively (for simplicity, we assume that I and J are finite). We will fix I and J throughout, and therefore suppress them in the notation. Define $X \equiv X_A \times X_B$ and $K \equiv I \cdot J$. To complete the specification of the environment, we must describe preferences. Let Γ be the set of functions v: $X \to R^2$ (mapping choices directly to felicities; we implicitly substitute for consumption as a function of choices). Note that since X is a finite set, we can think of Γ as the set of vectors in R^{2K}. For $\alpha_A, \alpha_B \in [0,1]$, utilities are given by

$$u_i(x;\alpha_i) \equiv (1-\alpha_i)v_i(x) + \alpha_i v_j(x), \quad i \neq j,$$

where $x = (x_A, x_B)$. We define $x_i^*(\alpha_i)$ as the choice vector that maximizes $u_i(x, \alpha_i)$.

The following result establishes that sufficiently high levels of altruism are almost always beneficial. The proof (omitted) is available from the authors on request.

THEOREM: *There exists a set* $\bar{\Gamma}$ *that is both open and dense in* Γ, *with the following property. Suppose* $v \in \bar{\Gamma}$. *Let* α_A^n *and* α_B^n *be sequences of numbers such that* $\alpha_A^n + \alpha_B^n < 1$, *and* $\lim_{n\to\infty}(\alpha_A^n + \alpha_B^n) = 1$. *Then there exists* N *such that for all* $n > N$ *there is a pure strategy Nash equilibrium* $\hat{x}^n \in X$, *and*

$$\lim_{n\to\infty} \left[u_i(x_i^*(\alpha_i^n), \alpha_i^n) - u_i(\hat{x}^n, \alpha_i^n)\right] = 0$$

for $i = A, B$.

This result says that when altruism is sufficiently strong (but not perfect), there is almost always an equilibrium in which payoffs are arbitrarily close to the global optima for both players. It is reassuring to note that the examples in Parts A and B both satisfy this property.

---

[11] We would like to thank James Andreoni for raising this point.

[12] An example is available from the authors upon request.

While high levels of altruism are generally beneficial, this should provide little comfort to those who subscribe to the traditional views mentioned at the outset of this paper. The effects of altruism on economic interaction are complex, and must be assessed on a case-by-case basis.

III. Conclusions

Our analysis is directly relevant to the literature on the economics of family behavior, in that it calls into question the views mentioned at the outset of this paper. One should not necessarily expect to find strong altruistic linkages between spouses, nor should altruism necessarily contribute to marital stability. Similar remarks apply to parent-child and sibling relationships, as well as to relationships outside the family that are based upon altruism or trust. Instead, our findings are consistent with pragmatic folk wisdom that distinguishes between love and compatibility.

These findings suggest potentially profitable directions for subsequent research on the family. Ideally, one should endogenize the degree of altruism within families by allowing for choice among potential spouses, alternative methods of child-rearing, and so forth. The key question is whether one would then expect the formation of families in which altruistically motivated resource transfers play a key role. If high levels of altruism fail to arise endogenously, then this would call into question the line of research advocated by Becker and Barro (among others), and would suggest that family decisions are more properly modeled as negotiations among primarily self-interested agents.

Our findings also have important implications in other contexts. We will mention two additional applications. First, when externalities generate inefficiencies (pollution, overpopulation), authorities typically resort to "moral suasion," regulation, or taxation. It seems to us that the first remedy is both most common and least studied. Yet the efficacy of this remedy depends upon the critical assumption that the development of a "social conscience" will produce more efficient outcomes. Our results suggest that this supposition is not of universal validity.

Second, biologists have pointed out that if individuals are the relevant evolutionary units, then evolution should be guided by selfish behavior. However, if the units of evolution are genes rather than individuals, then genes that incline organisms toward cooperation with relatives may prosper at the expense of less altruistic genes. Our analysis suggests that it may be important to refine this observation. Detailed study of particular environments could conceivably help biologists to better predict when and how altruism might evolve.[13]

Similar issues arise with respect to the evolution of social units (see, for example, the work of Axelrod, 1984, or Frank, 1986). Detailed study of evolutionary processes in which individuals bequeath altruistic preferences (rather than proclivities to cooperate, as in Axelrod's work) to their progeny and either survive or die off on the basis of realized felicity might well reveal that evolution tends to produce incomplete altruism.

---

[13] We would like to thank Theodore Bergstrom for suggesting this application.

REFERENCES

Abreu, Dilip, "Extremal Equilibria in Oligopolistic Supergames," *Journal of Economic Theory*, June 1986, *39*, 191–225.

Axelrod, Robert, *The Evolution of Cooperation*, New York: Basic Books, 1984.

Barro, Robert J., "Are Government Bonds Net Wealth?," *Journal of Political Economy*, November-December 1974, *81*, 1095–1117.

Becker, Gary S., "A Theory of Marriage: Part I," *Journal of Political Economy*, July-August 1973, *81*, 813–46.

______, (1974a), "A Theory of Social Interactions," *Journal of Political Economy*, November-December 1974, *82*, 1063–93.

______, (1974b), "A Theory of Marriage: Part II," *Journal of Political Economy*, March-April 1974, *82*, S11–S26.

______, *A Treatise on the Family*, Cambridge: Harvard University Press, 1981.

______, **Landes, Elisabeth M. and Michael, Robert T.**, "An Economic Analysis of Marital Instability," *Journal of Political Economy*, December 1977, *85*, 1141–87.

Bernheim, B. Douglas, Shleifer, Andrei and Summers, Lawrence, "The Manipulative Bequest Motive," *Journal of Political Economy*, December 1985, *93*, 1045–76.

Bergstrom, Theodore, "A Fresh Look at the Rotten Kid Theorem—and Other Household Mysteries," mimeo., University of Michigan, 1986.

Bruce, Neil and Waldman, Michael, "The Rotten Kid Theorem Meets the Samaritan's Dilemma," mimeo., UCLA 1986.

Frank, Robert H., "The Nature of the Utility Function," in A. MacFadyen and H. MacFadyen, eds., *Economic Psychology*, Amsterdam: North-Holland, 1986.

Friedman, James W., *Oligopoly and the Theory of Games*, Amsterdam: North-Holland, 1977.

Hirshleifer, Jack, "Shakespeare vs. Becker on Altruism: The Importance of Having the Last Word," *Journal of Economic Literature*, June 1977, *15*, 500–502.

Selten, Reinhard, "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory*, 1975, *4*, 22–55.

Implicit Contracts, Labor Mobility, and Unemployment

By RICHARD J. ARNOTT, ARTHUR J. HOSIOS, AND JOSEPH E. STIGLITZ*

When workers' search efforts are unobservable, the provision of insurance against firm-specific shocks adversely affects their incentives to find better jobs. In consequence, the equilibrium contract prescribes low wages and underemployment to encourage workers to leave low-productivity firms; and it employs both quits and layoffs to induce separations, with the mix depending both on the relative efficiency of on- and off-the-job search and on the search-incentive effects of layoffs.

Wage rigidities have long played a central role in Keynesian explanations of unemployment. Though a major aim of implicit contract theory was to provide an explanation for these rigidities (Costas Azariadis, 1975; Martin Baily, 1974), it has become increasingly apparent that these rigidities do not comprise a complete explanation of unemployment; that is, while the simpler versions of implicit contract theory successfully explain real wage rigidities as a consequence of risk-averse workers' demand for insurance against variations in the value of their marginal product (VMP), they cannot explain unemployment.[1]

To explain unemployment, one must explain, first, why reductions in demand induce layoffs rather than just reduced hours, and second, why those on layoff do not immediately secure employment elsewhere. Traditional implicit contract theory, by contrast, ignored both issues: it *assumed* that reductions in labor demand take the form of layoffs; and it *assumed* that all separated workers are immobile, due, say, to prohibitive mobility costs. The latter immobility assumption is particularly important since it implies that (*i*) laid-off workers are necessarily unemployed, (*ii*) quits are absent, (*iii*) high-demand firms cannot hire newly separated workers, and (*iv*) when optimally determined severance pay is provided, laid-off workers are better-off than retained workers. In fact, variable hours, layoffs, quits, and interfirm mobility are all prominent features of modern labor markets, and the prediction that workers prefer to be laid off is generally accepted as counterfactual.[2]

We propose here a model of implicit contracts which resolves most of these difficulties by relaxing the assumption that workers are immobile. We follow traditional contract theory in assuming that workers are risk-averse and firms are risk-neutral. As a result,

---

*Department of Economics, Boston College, Chestnut Hill, MA 02167 and Queen's University, Kingston, ON K7L 3N6, Canada; Department of Economics, University of Toronto, Toronto, ON M5S 1A1 Canada; Department of Economics, Stanford University, Stanford, CA 94305, respectively. We have benefited substantially from the comments of two anonymous referees. They alerted us to the issue of alternative sequencings of quits and layoffs and induced us to streamline the text. Arnott and Hosios would like to thank the Social Sciences and Humanities Research Council of Canada for financial support, while Arnott and Stiglitz would like to thank the National Science Foundation and the Olin Foundation.

[1] There are now a large number of surveys and critical reviews in this area, including George Akerlof and Hajime Miyazaki, 1980; Azariadis, 1979, Azariadis and Joseph Stiglitz, 1983; Oliver Hart, 1983; and most recently Sherwin Rosen, 1985.

[2] Though more recent developments in implicit contract theory, particularly those based on asymmetries of information (Azariadis, 1983; Sanford Grossman and Hart, 1981) have recognized the failure of the traditional versions to explain unemployment, they have failed to address the central problems raised above and have faced some further difficulties as well. For example, not only do these new theories obtain unemployment only under restrictive conditions but, when they do, they obtain it for all states of nature except the very best.

firms provide individuals with insurance. If labor mobility were costless and instantaneous, no insurance would be required against *relative* shocks (i.e., those which change the relative VMP's of labor in different uses) since each worker would move immediately to the job where his VMP is highest.

In fact, search is not only an individually costly activity, but is also a private activity that is costly to monitor. It is difficult for an employer to monitor his workers' search efforts and outside wage offers or even to verify whether his laid-off workers have become reemployed elsewhere. In turn, because the employer is unable to observe these search inputs and outcomes, the following moral hazard problems arise.

The first problem results when workers' job offers are private information, even when their search effort is observable. In this case, for each VMP realization, the firm can provide only a fixed severance payment to workers who quit or are laid off, and a given wage to workers who are retained. Consequently, the more insurance a firm provides against its own low-VMP draws (i.e., the greater the wedge between its workers' wage and VMP), the more likely it is that workers will refuse outside wage offers that exceed their current VMP but are less than their current wage, and hence the less efficient will be quit behavior.

The second moral hazard problem stems from the unobservability of search effort or intensity. In this case, the more insurance a firm provides against low-VMP draws, the less incentive a worker has to search for alternative employment and hence, once again, the less efficient will be quit behavior. Since it is privately (and socially) costly for individuals to remain where the value of their marginal product is low, the equilibrium insurance contract will in both situations provide incomplete insurance.

With private job-search information, and as a consequence of these moral hazard problems, the equilibrium implicit contract will be shown to prescribe relatively low wages and underemployment to encourage quits in bad times, and relatively high wages and overemployment to discourage quits in good times.[3] In this setting, moreover, quits (induced by relatively low wages) and layoffs are alternative means of creating separations from low-VMP firms.

On one hand, reduced insurance (lower wages for retained workers) has the potential advantage of encouraging quits by the lowest search cost workers, effectively discriminating among otherwise indistinguishable workers. On the other hand, layoffs have the potential twofold advantage of forcing workers to use a different and likely more efficient off-the-job search technology and, when quits precede layoffs, of inducing more intensive on-the-job search by decreasing the returns to those who do not quit. As these two instruments, reduced wage insurance and layoffs, are imperfect substitutes, we expect to observe both quits and layoffs at individual firms.

In a limiting (and we would argue unrealistic) version of our model, in which there is no search, we obtain the standard counterfactual results. We therefore contend that an analysis of the role of implicit contracts in understanding unemployment must focus on questions of labor mobility and private job-search information.[4] Regrettably, the simplest model that remedies the difficulties we have noted in the traditional framework must be somewhat complex: it must incorporate endogenous search; it must allow for the possibility of severance pay; and it must provide firms with a choice between layoffs and wage policy as means of encouraging the movement of workers from low- to high-VMP firms.

---

[3] In the relevant literature, under- (over-) employment is said to occur when a worker's shadow value of leisure is less (greater) than his value of marginal product. Note that asymmetric information plays a key role in our analysis, but that the particular source of informational asymmetry is different from that discussed, say, by Azariadis, 1983 or Grossman and Hart, 1981. In these papers workers cannot observe realizations of firms' revenue functions, which leads to an adverse selection problem. In the present paper, as in Jon Strand's, forthcoming, firms cannot observe workers' search efforts, which leads to a moral hazard problem.

[4] Related papers which also highlight labor mobility and asymmetric job offer information include those by John Geanakoplos and Takatoshi Ito, 1982; and Arthur Hosios, 1986, 1987; these papers develop simple general equilibrium models of layoffs, hiring and unemployment in which the *ex post* job-finding probability is endogenous, but quits are absent.

We begin in Section I with an analysis of firms' wages, severance pay, and internal distortions when workers' search efforts are private information and lay-off rates are given; Section I contrasts our results with earlier papers that assume either that separated workers are immobile or that job-search information is public. Section II presents a general model of layoffs and quits and compares implicit contracts that prescribe layoffs before and after on-the-job search; Section II also describes the equilibrium layoff rate in special cases with and without quits. Section III presents concluding remarks.

I. The Model

We develop a two-period model of a competitive economy which is buffeted by firm-specific shocks. During the *ex ante* period when there is still uncertainty, workers are free to join the firm whose employment contract offers the highest level of expected utility. At the beginning of the *ex post* period, after all random variables are realized, a fraction of each firm's initial labor pool may be laid off as prescribed by the contract. Subsequently, after search, some retained workers and some laid-off workers secure new employment elsewhere. Finally, production takes place at the end of the *ex post* period. See Figure 1.

While this formulation may initially appear restrictive, it will become clear as we proceed that this is not the only possible interpretation of our model (see fns. 10 and 11), and that most of our qualitative results, for instance concerning incomplete insurance, underemployment, and the presence of layoffs and quits together, would hold under alternative sequencings as well. In particular, in Section II we show that our results are not substantially altered when on-the-job search precedes layoffs. Until then, however, the formulation in Figure 1 will be used to simplify the exposition.

The critical informational assumption of the model is that each worker's *ex post* search effort and consequent job offers (if any) are privately known only by that worker.[5]

Firms are *ex ante* identical and risk-neutral, and labor is the only variable input to production. The VMP of a worker who supplies h units of labor is θh, where θ is the realization of a firm-specific shock; there are two equiprobable values of θ, θ_L, and θ_H with $\theta_L < \theta_H$, where $L(H)$ indexes the low- (high-) productivity state. Workers have identical tastes and technical ability. As in the traditional literature, they are assumed to have no consumption good endowment and savings are not allowed.

Firms offer employment contracts to attract employees during the *ex ante* period which are designed to maximize expected profits, and workers choose among these contracts to maximize expected utility. A typical contract will be denoted by $C = \{(w_L, h_L), r, s, (w_H, h_H)\}$. Under this contract employed workers are paid w_i for h_i units of labor when $\theta = \theta_i$; a worker is retained with probability r in the low-VMP state and hence laid off with probability $1-r$; all workers are retained in the high-VMP state; and each laid-off worker is given the fixed nonnegative severance payment s.[6]

---

[5] We also assume that each firm's technology is observable by the firm and its workers but not by others. This gives risk-neutral firms on informational advantage over external agents, so that job-related insurance is provided to workers by employers rather than insurance companies.

[6] Observe that asymmetric information has ruled out compensation conditional upon a worker's search effort or outside job offer.

We require severance pay to be nonnegative because of the enforcement (default) problems associated with negative severance pay; the implications of negative severance pay are, however, described later in Section II. Severance pay for workers who quit is prohibited here only to simplify the discussion; see Section I, Part D, and fn. 23.

This list of contract parameters is not exhaustive. For example, we could allow firms to reemploy laid-off workers who are unsuccessful in finding alternative employment. In this case, if layoffs are to have the desired off-the-job search incentive effects, firms will not reemploy workers at terms that make unsuccessful search a desirable outcome. Since the same moral hazard issues emphasized throughout this paper would arise here as well, our results do not depend substantially on such

PERIOD 1

Workers choose among contracts specifying:

h_L = hours to be worked in the low-productivity state;
h_H = hours to be worked in the high-productivity state;
w_L = payment to employed workers in the low-productivity state;
w_H = payment to employed workers in the high-productivity state;
r = probability of being retained in the low-productivity state;
s = severance payment to laid-off workers.

PERIOD 2

Realization of labor productivity:

$\theta \varepsilon \{\theta_L, \theta_H\} \quad \theta_L < \theta_H.$

Firms make their employment decisions:

if $\theta = \theta_H$, worker is retained with certainty;
if $\theta = \theta_L$, worker is retained with probability r and laid off with probability 1-r.

Workers make their search decisions:

retained workers choose $e(w_i, h_i)$ when $\theta = \theta_i$;
laid-off workers choose e(s).

Searching workers receive job offers:

retained workers who < accept their best offer quit to work elsewhere;
retained workers who < reject their best offer remain employed at their initial firm;
laid-off workers who < accept their best offer become re-employed elsewhere;
laid-off workers who < reject their best offer, or who receive no offers, remain unemployed.

Production takes place.

FIGURE 1

The terms of employment offered by firms in the *ex ante* period will depend on what alternative opportunities are available to workers in the *ex post* period. We assume that there is a spot labor market characterized by a wage distribution, but in which all jobs require the same number of hours $\bar{h} > 0$.[7] Workers are concerned with the *highest* wage they can find, and this in turn depends on their search intensity and whether they are employed while searching.

We let $F(z,e)$ and $G(z,e)$ denote the distribution functions for the highest wage offers, z, given search intensity e, for those workers who are employed and unemployed, respectively; f and g are the corresponding density functions. F and G are assumed to satisfy: $F(0,0) = G(0,0) = 1$; and $F_e, G_e < 0$ for $z > 0$. In words, a worker cannot effortlessly receive an offer of employment; and the probability that his best offer will exceed z is an increasing function of his search intensity.

Taking their employment status as given, workers choose the level of search effort to maximize their expected utility. After substi-

---

terms of reemployment. We have therefore chosen to omit them.

[7] The assumption that all alternative jobs require the same number of hours could be removed at some increase in notational complexity. We again emphasize that we are concerned with constructing the simplest model illustrating the points at issue, rather than the most complete.

tution, the resulting maximum expected utility enjoyed by retained workers is simply a function of their wages and hours, $U(w, h)$; similarly, the maximum expected utility attained by laid-off workers is a function only of their severance pay, $V(s)$. Thus, dropping the common multiplicative factor 1/2, total expected utility from contract C is[8]

$$W(C) = rU(w_L, h_L) + (1-r)V(s) + U(w_H, h_H).$$

Taking account of workers' search and quit behavior, firms' profits per retained employee are a function of the state θ and of the worker's wages and hours, $\pi(w, h, \theta)$, whereas the cost per laid-off worker is simply the severance payment, s. Thus, total expected profit per employee is given by

$$Z(C) = r\pi(w_L, h_L, \theta_L) - (1-r)s + \pi(w_H, h_H, \theta_H).$$

As workers are freely mobile in the *ex ante* market, competition among firms will drive expected profits to zero. Hence the equilibrium contract is found by solving

$$\text{(P)} \qquad \max W(C) \text{ s.t. } Z(C) = 0.$$

The remainder of this section describes and interprets the first-order conditions of this problem. Among the questions we ask are: Under what conditions will the wage in the bad state be below that in the good state, that is, when will there be at least partial wage flexibility? When will those who stay with the firm be underemployed? And finally, if there are layoffs, when will laid-off workers be worse-off than retained workers? Until Section II, the retainment probability, r, is taken as given.

A. *Derivation of Expressions for Expected Utility*

To answer these questions, we must first derive explicit expressions for a worker's expected utility when he is retained and when he is laid off. A worker's instantaneous utility function is given by $\alpha(w, h)$. When a worker is retained, his expected utility, $u(w, h, e)$, is the utility from his current job, $\alpha(w, h)$, times the probability he stays with his current employer (i.e., does not get a better offer), plus the expected utility if he quits, $\sigma(x, e)$, times the probability of quitting, minus the disutility, $\beta(e, h)$, of searching for a better job at intensity e:[9]

$$u(w, h, e) = \alpha(w, h)F(x, e) + \sigma(x, e) \times (1 - F(x, e)) - \beta(e, h),$$

where $x = x(w, h)$ is the wage offer which makes the worker indifferent between quitting and staying, and is defined implicitly by $\alpha(w, h) = \alpha(x, \bar{h})$, $F(x, e)$ is the probability that a worker's highest wage offer is less than x given on-the-job search intensity e, and the expected utility of a worker who quits is

$$\sigma(x, e) = \left(\int_x^\infty \alpha(z, \bar{h}) f(z, e)\, dz \right) \Big/ (1 - F(x, e)).$$

We assume that α is a strictly concave function satisfying $\alpha_w > 0$ and $\alpha_h < 0$ and that β is an increasing convex function of e satisfying $\beta_h \geq 0$ and $\beta(0, h) = 0$.[10,11] Severance

---

[8] Workers are assumed not to quit to search, that is, $U(w_i, h_i) \geq V(0)$. This condition is assumed to be satisfied, and will be ignored in the subsequent discussion.

[9] The additively separable structure and the imposition of $\beta_w = 0$ simplify the analysis but are otherwise inessential.

[10] The assumption, $\beta_h \geq 0$, is that an increase in hours worked cannot decrease the disutility of search effort. More than one interpretation of the model is possible and each involves a slightly different rationale for this assumption.

Situations where an increase in hours worked leaves the worker with less time for search (or other nonwork) activities, obviously come to mind. One such interpretapay for retained workers who quit is considered later in Section I, Part D.

Analogously, the expected utility of a laid-off worker who receives severance pay s and expends e on search equals

$$v(s,e)=\alpha(s,0)G(y,e)$$
$$+\int_{y}^{\infty}\alpha(z+s,\bar{h})g(z,e)\,dz-\beta(e,0),$$

where $y=y(s)$ is the lowest acceptable wage offer and is defined by $\alpha(s,0)=\alpha(y+s,\bar{h})$. In the absence of search, observe that $u(w,h,0)=\alpha(w,h)$ and $v(s,0)=\alpha(s,0)$.

Retained and laid-off workers' optimal on- and off-the-job search intensities are described, respectively, by $e(w,h)=\operatorname{argmax}u(w,h,e)$ and $e(s)=\operatorname{argmax}v(s,e)$. Hence their corresponding equilibrium utility levels are $U(w,h)=u(w,h,e(w,h))$ and $V(s)=v(s,e(s))$; and firms' equilibrium quit and profit functions per retained worker are, respectively,

$$q(w,h)=1-F(x(w,h),e(w,h)),$$
$$\pi(w,h,\theta)=(1-q(w,h))(\theta h-w).$$

The quit rate is a function only of the wages and hours offered on the job and equals the probability of funding a higher utility job given the optimally chosen search effort, $e(w,h)$. Similarly, the job-finding rate for laid-off workers, $1-G(y(s),e(s))$, is a function only of the severance payment s.

B. *Wages and Production Efficiency*

In this section, we shall show how firms' attempts to encourage mobility in the low-VMP state leads to incomplete insurance and, under normal conditions, to underemployment in this state.

From the first-order conditions of problem (P), with respect to wages and hours, we obtain

$$\text{(1a)}\quad \frac{U_w(w_L,h_L)}{U_w(w_H,h_H)}=\frac{\pi_w(w_L,h_L,\theta_L)}{\pi_w(w_H,h_H,\theta_H)},$$

$$\text{(1b)}\quad \frac{U_h(w_i,h_i)}{U_w(w_i,h_i)}=\frac{\pi_h(w_i,h_i,\theta_i)}{\pi_w(w_i,h_i,\theta_i)},$$
$$i=L,H.$$

The solution to (P) thus equates agents' marginal rates of substitution between wages across states, and between wages and hours within each state. Observe that these expressions measure the direct *plus* incentives-related substitution possibilities among the terms of employment for workers and firms;

---

tion of the model, where search and production occur together during an initial subperiod and where hours worked directly affect search disutility, is formalized below in fn. 11.

Somewhat different interpretations are also possible. For example, suppose that each period is divided into two parts, a job-search period and a production period, and that a worker's current employer is required to tell him at the beginning of the job-search period whether he will be retained for the production period. Contract design will then have to account for the effect of this announcement on workers' behavior during the initial search period. In this case, with search preceding production, it is the substitutability of leisure during the search and production subperiods that results in $\beta_h\geq 0$.

[11] We now formulate a slightly different interpretation of the events to show that whether search precedes or is concurrent with production is not critical here.

Suppose the unit period is divided into 2 subperiods, of lengths t and $1-t$, such that workers supply labor and search during the first subperiod, and either quit or stay at the beginning of the second subperiod, and therefore

$$u(w,h,e)=t[\alpha(w,h)-\beta(e,h)]+(1-t)$$
$$\times[\alpha(w,h)F(x,e)+\sigma(x,e)(1-F(x,e))].$$

Then, defining $\hat{F}=t+(1-t)F$ and $\hat{\beta}=t\beta$, we get

$$u(w,h,e)=\alpha\hat{F}+\sigma(1-\hat{F})-\hat{\beta},\qquad \text{where } \hat{F}(0,0)=1.$$

That is, except for the circumflex "$\hat{}$," this is identical to the expression for $u(w,h,e)$ in the text, and hence our analysis of search and quits goes through in exactly the same way. Allowing different wage-hours packages in each subperiod would complicate matters without changing our main results—the provision of partial insurance and the possibility of under- or overemployment with asymmetric search information.

and that the latter indirect effects take into account the adjustment of workers' search efforts and quit rates.

Returning to the basic equations defining U and π, we can derive, for $i = L, H$,

$$\text{(2a)} \quad U_w(w_i, h_i) = (1 - q(w_i, h_i))\alpha_w(w_i, h_i),$$

$$\text{(2b)} \quad U_h(w_i, h_i) = (1 - q(w_i, h_i))\alpha_h(w_i, h_i) - \beta_h(e(w_i, h_i), h_i),$$

$$\text{(2c)} \quad \pi_w(w_i, h_i, \theta_i) = -(1 - q(w_i, h_i)) - A_i q_w(w_i, h_i),$$

$$\text{(2d)} \quad \pi_h(w_i, h_i, \theta_i) = (1 - q(w_i, h_i))\theta_i - A_i q_h(w_i, h_i),$$

where

$$A_i = \theta_i h_i - w_i.$$

Notice that A_H and $-A_L$ are respectively equal to workers' implicit insurance premium and indemnity.

Immobile Labor.[12] In this case our model generates the standard results of full insurance and production efficiency; that is, substituting $q(w_i, h_i) = 0$ into (2), (1a,b) respectively simplify to

$$\alpha_w(w_L, h_L) = \alpha_w(w_H, h_H),$$

$$\frac{-U_h(w_i, h_i)}{U_w(w_i, h_i)} = \frac{-\alpha_h(w_i, h_i)}{\alpha_w(w_i, h_i)} = \theta_i, \qquad i = L, H.$$

As expected, the equilibrium contract equates workers' marginal utility of income across states and equates their marginal rate of substitution and marginal product in each state. Also, if either α is additively separable or the supply of labor is inelastic ($h_L = h_H$), we then get the well-known rigid contract wage result, $w_L = w_H$.

Mobile Labor: *Incomplete Insurance*. Suppose workers' equilibrium quit rates are nonzero in both states. In this case, a higher contract wage reduces the expected return to on-the-job search and, as a result, retained workers' search effort and consequent quit rate fall when the wage is raised, that is, $q_w(w_i, h_i) < 0$.[13] Thus, making use of the fact that $A_L < 0 < A_H$, (1a) and (2a,c) give

$$\text{(3)} \qquad \alpha_w(w_L, h_L) > \alpha_w(w_H, h_H).$$

That is, the marginal utility of income is higher in the low-VMP state than in the high-VMP state and, therefore, *the optimal contract no longer provides complete insurance*.

Since workers who have searched unsuccessfully are indistinguishable from those who make no attempt to search at all, firms will in part use the contract wage as an incentive device. Thus, when firms are subsidizing retained workers in the low-VMP state as part of the insurance package, w_L is lowered (relative to the complete insurance situation) to encourage them to quit; and when workers are being taxed in the high-VMP state, w_H is raised to discourage quits. In particular, if the utility function α is separable, or labor supply is inelastic, (3) gives $w_H > w_L$.[14]

---

[12] Workers are immobile when they simply choose to forego search, setting $e(w,h) = 0$ and hence $q(w,h) = 0$; this outcome is most likely when search costs are high and there are few well-paying jobs in the spot market.

[13] $q = 1 - F(x,e)$ gives $q_w = -(F_x x_w + F_e e_w)$, where $F_x > 0$ and $F_e > 0$. From the definition of $x(w,h)$, $x_w = \alpha_w(w,h)/\alpha_w(x,\bar{h}) > 0$; and from the f.o.c. $u_e(w,h,e) = 0$, where

$$u_e = \alpha(w,h)F_e(x,e) + \int_x^\infty \alpha(z,\bar{h})f_e(z,e)\,dz - \beta_e(e,h)$$

and $x = x(w,h)$, we have $e_w = -\alpha_w(w,h)F_e/u_{ee}$. Then $q_w = -\alpha_w B$ where $B \equiv (F_x/\alpha_x) - F_e^2/u_{ee} > 0$ when the second-order condition, $u_{ee} < 0$, is satisfied.

[14] Previous work has recognized that wages will likely be raised to discourage quits when profits are high or turnover is costly (for example, Steve Salop, 1979), but has failed to make the symmetrical argument when profits are low. Separability is sufficient but not necessary for a positive correlation between real wages and VMP's.

Mobile Labor: *Underemployment*. We now examine whether our model generates underemployment. In the present search context, however, it is not immediately clear how this term should be defined. In particular, is there underemployment when $-U_h/U_w < \theta$, that is, when the marginal rate of substitution between wages and hours is less than the marginal rate of transformation, taking into account induced search; or is there underemployment when $-\alpha_h/\alpha_w < \theta$, that is, when the MRS between wages and hours along the instantaneous utility function is less than the MRT?

Since it is easier to establish results concerning the relative magnitudes of $-U_h/U_w$ and θ, we henceforth use the terms under- and overemployment to describe situations, where $-U_h/U_w <$ and $> \theta$, respectively, and say production efficiency obtains when the two are equal.[15]

When quit rates are nonzero, (1b) and (2c, d) give

$$(4)\quad \frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} = \frac{\theta_i - A_i q_h(w_i,h_i)/(1-q(w_i,h_i))}{1+A_i q_w(w_i,h_i)/(1-q(w_i,h_i))}.$$

Earlier we noted that increasing wages reduces quit rates. If increasing hours also reduce quit rates, so that $q_h < 0$, we get

$$(5)\quad \frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} \gtreqless \theta_i \quad \text{as} \quad A_i \gtreqless 0.$$

Therefore, *when increasing hours worked reduce quits, the equilibrium contract prescribes underemployment* (*to encourage quits*) *in those states when workers are being subsidized, and overemployment* (*to discourage quits*) *when profits are strictly positive*.[16,17]

The property $q_h < 0$ is sufficient but not necessary for (5). A necessary and sufficient condition for (5) is that the *compensated* quit derivative,

$$\left.\frac{dq}{dh}\right|_U = q_h + q_w \left.\frac{dw}{dh}\right|_U,$$

is negative; that is, *there will be underemployment* (*overemployment*) *in those states where profits are negative* (*positive*) *if and only if workers' compensated quit derivative with respect to hours is negative.*

To see this, we consider a perturbation to $\{w_i, h_i\}$ which keeps workers' expected utility in state i fixed and calculate the effect on firm profits. At the optimum

$$\left.\frac{d\pi}{dh}\right|_U = (1-q)\left[\theta - \left.\frac{dw}{dh}\right|_U\right] - (\theta h - w)\left.\frac{dq}{dh}\right|_U = 0.$$

Recalling that

$$\left.\frac{dw}{dh}\right|_U = -U_h/U_w,$$

and substituting, we can obtain

$$(6)\quad \frac{-U_h(w_i,h_i)}{U_w(w_i,h_i)} = \theta_i - \frac{A_i}{(1-q)}\left.\frac{dq}{dh}\right|_U,$$

yielding the stated result.

---

[15] From (2a, b), $U_h/U_w = \alpha_h/\alpha_w - \beta_h/[(1-q)\alpha_w]$, and hence the two possible definitions of underemployment coincide if $\beta_h = 0$, that is, if increasing hours worked does not affect the cost of searching (at any level of search effort). Normally, however, $\beta_h > 0$ (see fn. 10) in which case $-U_h/U_w > -\alpha_h/\alpha_w$. Accordingly, when there is underemployment under the first criterion (taking into account induced search), there is underemployment under the second criterion; and when there is overemployment under the second criterion (i.e., $-\alpha_h/\alpha_w > \theta$), there is overemployment under the first criterion.

[16] Analogous results concerning the relative magnitudes of α_h/α_w and θ follow straightforwardly with some obvious caveats (see fn. 15) and are omitted.

[17] These results are plausible yet appear to require $q_h < 0$. It is possible, however, that q_h can be positive, even for well-behaved utility functions. There are two opposing effects at work (see the expression for q_h in fn. 18). On one hand, an increase in hours worked (at fixed wages) makes the job less attractive, and this increases quit rates. On the other hand, if search is time-intensive, then a worker who works more has less time to search, and this reduces quit rates.

In turn, it can be shown that[18]

$$\left.\frac{dq}{dh}\right|_U = \frac{-B\beta_h}{(1-q)} - \frac{F_e\beta_{eh}}{u_{ee}}, \tag{7}$$

where $B > 0$ (see fn. 13), $F_e < 0$ and $u_{ee} < 0$. Therefore, *a sufficient condition in our model for the compensated quit derivative to be negative is that both β_h and β_{eh} be positive, that is, that increasing hours worked increases both the total and marginal costs of search.* Observe that the compensated quit derivative is negative under less restrictive conditions than are required for $q_h < 0$.[19] Observe also that the equilibrium contract satisfies $-U_h/U_w = \theta$ and hence exhibits production efficiency whenever the compensated quit derivative is zero.

The following two properties of these results are noteworthy. First, the occurrence of under- versus overemployment in our model does not depend on whether firms are risk-averse or, with reference to either $u(w, h, e(w, h))$ or $\alpha(w, h)$, whether leisure is a normal good, or finally, whether $\alpha(w, h)$ has a specific functional form, such as $\alpha = \mu(w) - \gamma(h)$ or $\alpha = \mu(w - \gamma(h))$.[20] Whether there is under- or overemployment depends instead only upon certain weak properties of the utility function (for example, $\beta_{eh} > 0$) and upon the contractual availability of insurance for workers ($A_i \neq 0$).

Second, (5)–(7) indicate that under reasonable conditions underemployment will occur only in states where workers are being subsidized, with the amount of underemployment positively related to the size of the subsidy; and that overemployment will be reserved for profitable states, with the amount of overemployment again responding to the level of profits. Thus, as seems reasonable, underemployment occurs in bad times and overemployment in good times.[21]

C. *On Severance Pay*

Severance pay is oftentimes portrayed as the Achilles' heel of implicit contract theory. The original models largely ignored severance pay. Later work introduced severance pay but obtained the counterfactual result that workers prefer to be laid off rather than retained; that is, contracts which equate (expected) marginal utilities of income across all states result in higher (expected) utilities among those on layoff, in which case one would expect contracts to contain reverse seniority layoff clauses. In this section we show, on the contrary, that under reasonable conditions concerning labor mobility and preferences, laid-off workers are worse off than retained workers. Before presenting our results, we review the standard ones.

The first-order conditions of problem (P) yield

$$-U_w(w_i, h_i)/\pi_w(w_i, h_i, \theta_i) = V_s(s) \tag{8}$$

whenever the optimal severance payment s is strictly positive. (Later, we identify cases where $s = 0$.) Observe that when both on- and off-the-job search occur, (1), (2), and (8)

---

[18] It can be shown that $q_h = -\alpha_h B + (F_e\beta_{eh}/u_{ee})$ where, as in fn. 13, $B = -q_w/\alpha_w$. Also, from (2a, b),

$$\frac{-U_h}{U_w} = \frac{-\alpha_h}{\alpha_w} + \frac{\beta_h}{\alpha_w(1-q)} = \left.\frac{dw}{dh}\right|_U,$$

and therefore, substituting $q_w = -\alpha_w B$ and the above expressions for q_h and $-U_h/U_w$ into

$$\left.\frac{dq}{dh}\right|_U = q_h + q_w\left.\frac{dw}{dh}\right|_U$$

gives (7).

[19] Since $q_h = -\alpha_h B - F_e\beta_{eh}/u_{ee}$, where $\alpha_h < 0$, (7) is negative under less restrictive conditions than are required for $q_h < 0$.

[20] See Russell Cooper, 1983; Hart, 1983; and Stiglitz, 1986, for a description of the critical role of preferences in generating production inefficiencies in models where the firm, but not the worker, can observe the state of nature.

[21] By contrast, earlier models generated either *under*-employment in all *high*-VMP states except the highest (Azariadis, 1983; Grossman and Hart, 1981) or *over*-employment in all *low*-VMP states except the lowest (V. V. Chari, 1983; Cooper, 1983; Jerry Green and Charles Kahn, 1983).

give

$$\alpha_w(w_L, h_L) > V_s(s) > \alpha_w(w_H, h_H) \quad \text{for } A_L < 0 < A_H,$$

which is the obvious extension of our earlier partial insurance result.

Reverse Seniority. There are two versions of the reverse seniority result. First, suppose employed and laid-off workers are both immobile (do not search) so that (8) becomes

$$\alpha_w(w_L, h_L) = \alpha_s(s,0). \tag{9}$$

The optimal (insurance) contract equates the marginal utilities of income of retained and laid-off workers. Differentiating $\alpha_w =$ constant, gives $dw/dh = -\alpha_{wh}/\alpha_{ww}$, and hence $d\alpha/dh = \alpha_h - \alpha_w\alpha_{wh}/\alpha_{ww}$ for fixed α_w. Therefore, (9) implies that

$$\alpha(w_L, h_L) < (>) \alpha(s,0),$$

when leisure is a normal (inferior) good; and hence *immobile workers prefer to be laid off when leisure is normal* (Cooper, 1983; Stiglitz, 1986).

Second, suppose employed workers are again immobile, and that laid-off workers are now mobile but face uncertain job opportunities. To simplify, suppose the utility function is separable, $\alpha(w,h) = \mu(w) - \gamma(h)$. In this case, (8) again yields the result that retained and laid-off workers' (expected) marginal utilities of income are equal, that is,

$$\mu'(w_L) = \mu'(s)G(y,e) + \int_y^\infty \mu'(s+z)g(z,e)\,dz, \tag{10}$$

where $e = e(s)$. Equation (10) implies that $\mu(w_L)$ will be greater than (equal to, less than)

$$\mu(s)G(y,e) + \int_y^\infty \mu(s+z)g(z,e)\,dz,$$

as $\mu(\ )$ is an increasing (constant, decreasing) absolute-risk-aversion utility function.[22] Thus, *whenever search entails zero disutility and hours worked is not a variable, workers with decreasing absolute risk-aversion utility functions prefer layoffs with severance pay.*

Costly Search. With costly search, however, the presumption that laid-off workers are better-off no longer obtains. There are two reasons for this. First, if search costs are not "completely pecuniary" (i.e., do not enter the utility function only additively with income), then providing full-income insurance (by equating expected marginal utilities) will not fully compensate for the costs and risks of search. Second, retained workers in low-VMP firms will, we have argued, typically have their hours reduced to encourage them to search more; as a result, hours may be reduced below the "standard" hours at alternative employment so that, even with full-income insurance, retained workers' expected utility may be higher than that of laid-off workers.

To see these results, first suppose every laid-off worker can effortlessly (with $e = 0$) secure a new job which pays z for $\bar{h}$; since there are no incentive problems here, the optimal severance payment will satisfy

$$\alpha_w(w_L, h_L) = \alpha_s(s+z, \bar{h}). \tag{11a}$$

It then follows that laid-off workers will be worse-off when leisure is normal whenever the spot job entails more work ($\bar{h} > h_L$). Now, suppose instead that this alternative job can be secured only by exerting some fixed effort $\hat{e} > 0$; assuming that laid-off workers choose to search, retained workers are better-off when

$$\alpha(w_L, h_L) > \alpha(s+z, \bar{h}) - \beta(\hat{e},0). \tag{11b}$$

---

[22] Whether the expected utility of income is greater or less for laid-off than for retained workers depends upon whether utility is a concave or convex function of marginal utility. Recognizing that $s+z$ is a mean marginal utility preserving spread of w_L, this result follows as a corollary to Peter Diamond and Stiglitz, 1974. See also Haruo Imai, John Geanakoplos, and Ito Takatoshi, 1981.

Whenever search costs are in part nonpecuniary, it is clear that (11a) and (11b) can both be satisfied under a wide range of conditions.

Thus, in the special case where employed workers are immobile we have the following result: Provided *i*) search costs are nonpecuniary (and accordingly, severance payments do not fully cover them), *ii*) leisure is normal, and *iii*) there are significant probabilities of being reemployed elsewhere at jobs requiring at least as much work as at their *ex ante* firms, laid-off workers will be worse-off than retained workers; and this will be true even in the presence of decreasing absolute risk aversion.

More generally, when the disutility of search is not an additively separable component of utility and retained workers also search on the job, it is clear that one can infer little from (8) about the relative magnitudes of $U(w_L, h_L)$ and $V(s)$.

D. *Severance Pay for Retained Workers*

Our results are not substantially altered if firms provide severance pay for retained workers who quit in bad states. Such severance pay would be used to encourage quits. However, so long as reducing hours worked reduces the cost of on-the-job search, underemployment will still occur in low-VMP states; so long as reducing wages increases the return to search, firms will not provide complete income insurance; so long as search costs are not completely additive with income, the severance pay will not fully compensate for effort expended on search; and finally, so long as there is imperfect information about job offers, the distortions we described earlier with respect to quit behavior will remain.[23]

II. Endogenous Layoffs

The previous section established that, with an exogenous layoff rate, the employment contract will underemploy workers in the bad state and provide only partial insurance to encourage workers to quit. In this section we shall argue that, under reasonable conditions, low-productivity firms will in fact choose to lay off some but not all of their workers.

Since workers in our model are technologically identical, and since hours per worker and the number of workers are perfect substitutes in production, differential productivity and technological arguments for layoffs have been ruled out.[24] In this section we show that the layoff rate will tend to be high when off-the-job search is relatively more efficient than on-the-job search and that quits will be the preferred mode of separation when on-the-job search is effective. These results are robust; they hold whether or not

---

[23] Suppose firms pay b to workers who quit. In this case, only those workers who receive an outside offer $z \geq x - b$ will quit, where $x = x(w,h)$ is again defined by $\alpha(w,h) = \alpha(x,\bar{h})$. As a result, $u(w,h,e)$ is replaced by

$$u(w,h,e,b) = \alpha(w,h)F(x-b,e) + \sigma(x,e,b) \times (1-F(x-b,e)) - \beta(e,h),$$

where

$$\sigma(x,e,b) = \int_{x-b} \alpha(z+b,\bar{h})f(z,e)\,dz \Big/ (1-F(x-b,e)).$$

Hence the optimal search intensity and resulting quit rate are given by

$$e(w,h,b) = \operatorname{argmax} u(w,h,e,b),$$

$$q(w,h,b) = 1 - F(x-b,e(w,h,b)),$$

where $q_b > 0$. Similarly, the resulting utility and profit functions are given by

$$U(w,h,b) = u(w,h,e(w,h,b),b),$$

$$\pi(w,h,b) = (1-q(w,h,b))(\theta h - w) - q(w,h,b)b.$$

As U_w and U_h are again described by (2a,b), and π_w and π_h by (2c,d), where $A = \theta h - w + b$, our earlier results on wages and production inefficiencies go through exactly as before. See Ito, 1986; and Charles Kahn, 1985, for further results on quits and severance pay.

[24] See Ken Chan and Yannis Ioannides, 1982; Mark Lowenstein, 1983; Rosen, 1985; Andrew Weiss, 1980; and Stiglitz, 1986.

the firm provides severance pay and whether layoffs precede or follow quits.

We begin by formulating a general model using the structure and notation of the preceding section; this model allows us to identify the principal determinants of the equilibrium layoff rate and to verify that whether layoffs precede or follow on-the-job search is not a critical factor. Section II, Part B particularizes the model to derive specific results.

A. *A General Formulation*

Suppose the *ex post* period is divided into two subperiods, of lengths t and $1-t$, such that workers can search and work in both subperiods in exactly the same manner as described earlier in Section I. Layoffs, meanwhile, can occur only once, at the beginning of either the first or second subperiod. We consider two cases; (all) layoffs precede quits (LQ) when layoffs occur at the beginning of the first subperiod (see Figure 2(a)) and (some) quits precede layoffs (QL) when layoffs occur at the beginning of the second subperiod (see Figure 2(b)). Section I considered the special case LQ with $t=1$.

Earlier we let $U = F\alpha + (1-F)\sigma - \beta$ denote the expected utility of a retained worker where $1-F$ is the quit rate that results with the optimal search effort. Now, using superscripts to denote subperiods $(i=1,2)$ and subscripts to denote states $(j = L, H)$, we adapt this notation and the corresponding expression for laid-off workers as follows. Let tU_j^1 and $(1-t)U_j^2$, where $U_j^i = \alpha_j^i F_j^i + \sigma_j^i(1-F_j^i) - \beta_j^i$, respectively, denote the expected utilities of a retained worker in state j and in subperiods 1 and 2; and let tV^1 and $(1-t)V^2$, defined analogously, respectively denote the expected utilities of a laid-off worker in subperiods 1 and 2.

When *layoffs precede quits*, total expected utility is

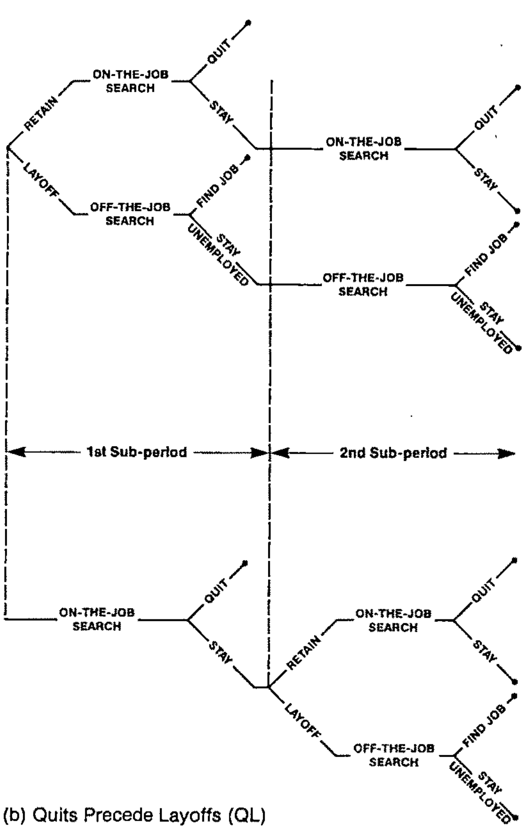
$$W^{LQ} = r\hat{U}_L + (1-r)\hat{V} + \hat{U}_H, \tag{12a}$$

where $\hat{U}_L$, $\hat{V}$, and $\hat{U}_H$ are the expected utilities of a worker conditional on his being retained in the low-VMP state, laid off in the low-VMP state, and retained in the high-VMP state, respectively, at the beginning of the first subperiod. Employing the notation introduced above, we have

$$\hat{U}_j = tU_j^1 + (1-t)\left[F_j^1U_j^2 + \left(1-F_j^1\right)\sigma_j^1\right] \qquad j = L, H,$$

$$\hat{V} = tV^1 + (1-t)\left[G^1V^2 + (1-G^1)\sigma^1\right].$$

For example, the expected utility of a worker retained at a low-VMP firm equals his expected utility in the first subperiod, tU_L^1, *plus* the expected utility in the second subperiod conditional on having quit during the first, $(1-t)\sigma_L^1$, times the probability of quitting during the first subperiod, $(1-F_L^1)$, *plus* the expected utility in the second subperiod conditional on having not quit during the



(a) Layoffs Precede Quits (LQ)

(b) Quits Precede Layoffs (QL)

FIGURE 2

first, $(1-t)U_L^2$, times the probability of staying throughout the first subperiod, F_L^1. Total expected profit per worker is

$$(12b) \quad Z^{LQ} = r\hat{\pi}_L - (1-r)s + \hat{\pi}_H,$$

where $\hat{\pi}_j = t\pi_j^1 + (1-t)F_j^1\pi_j^2$, and $\pi_j^i = F_j^i(\theta_j h_j^i - w_j^i)$.

When the layoff decision is made at the beginning of the second subperiod, so that *quits precede layoffs*, total expected utility is

$$(13a) \quad W^{QL} = tU_L^1 + (1-t) \times \left\{ F_L^1\left[rU_L^2 + (1-r)V^2\right] + \left(1-F_L^1\right)\sigma_L^1 \right\} + \hat{U}_H.$$

Observe that all workers at low-VMP firms are retained during the first subperiod and that from among those who stay to the second subperiod, a fraction $(1-r)$ are laid off. Total expected profit per worker is

$$(13b) \quad Z^{QL} = t\pi_L^1 + (1-t) \times F_L^1\left[r\pi_L^2 - (1-r)s\right] + \hat{\pi}_H.$$

The optimal contract when layoffs precede quits is found by maximizing (12a) subject to the constraint that (12b) is zero; and the optimal contract when quits precede layoffs is found by maximizing (13a) subject to the constraint that (13b) is zero. Whether the equilibrium contract prescribes layoffs before or after on-the-job search is determined by comparing the solutions to these two problems. Fortunately, as shown below, it is not necessary to make this comparison to determine whether layoff rates are ever positive.

When layoffs are absent, so that the retainment probability equals one, the payoffs for workers and firms are the same whether layoffs precede or follow quits; that is, at $r=1$, (12a) and (13a) are the same and (12b) and (13b) are the same. Therefore, if the optimal layoff rate is positive for one sequence of events and is zero for the other, *workers must strictly prefer the sequence that delivers the positive layoff rate*. In particular, therefore, the optimal contract when layoffs precede quits strictly dominates the optimal contract when quits precede layoffs whenever the former prescribes layoffs but the latter does not.

Although the determinants of layoffs when layoffs precede and follow quits are different, sufficient conditions for $r<1$ under either sequence are also sufficient when the sequence is itself a choice variable. As a result, we can investigate each case separately.

When *layoffs precede quits*, the derivative of the Lagrangian corresponding to (12), $L^{LQ} = W^{LQ} + \lambda Z^{LQ}$, with respect to the retainment rate is

$$(14a) \quad \frac{\partial L^{LQ}}{\partial r} = \hat{U}_L - \hat{V} + \lambda(\hat{\pi}_L + s),$$

where λ is the multiplier corresponding to the zero-expected profit constraint and, from the other first-order conditions, is interpreted as the marginal utility of income. Retaining an extra worker has two effects: it makes that worker better off by $\hat{U}_L - \hat{V}$; and it increases the firm's subsidization costs in the low-productivity state by $-(\hat{\pi}_L + s)$ which, since this cost is passed on, lowers the expected utility of all workers.

Thus, taking the other contract parameters as given, we expect more efficient *on-the-job* search to increase the benefit of retaining an extra worker by raising the expected utility on the job, $\hat{U}_L$, and to decrease the cost of retaining an extra worker by inducing more quits and hence lowering the expected subsidy, $-\hat{\pi}_L$. By contrast, we expect more efficient *off-the-job* search to decrease the benefit of retaining an extra worker by raising the expected utility off the job, $\hat{V}$. Therefore, as off-the-job search becomes relatively more efficient, we expect that the equilibrium mix of quits and layoffs will shift from mostly quits to mostly layoffs.[25] Section II, Part B and the Appendix, (Part A) derive these results using a version of the above model that entails a specific formalization of the notion of "more efficient search."

---

[25] This shift will be continuous and interior solutions $(0<r<1)$ will result if workers' total expected utility is strictly concave in the retainment rate, *mutatis mutandis*; this function will normally be strictly concave due to the strict concavity of the underlying instantaneous utility function, $\alpha(w,h)$ (see Section II, Part B).

When *quits precede layoffs*, there are three factors at work determining the layoff rate: First, there is the insurance-mobility-cost tradeoff described above that is the sole determinant when layoffs precede quits. Second, there is the *screening effect* of on-the-job search which tends to decrease the layoff rate. When workers differ in their search ability, allowing on-the-job search to precede layoffs screens out the efficient searchers and leaves the least mobile workers behind. In effect, this initial screening makes subsequent off-the-job search less efficient, on average, which in turn makes subsequent layoffs less attractive. Finally, there is the *incentive effect* of layoffs which tends to increase the layoff rate. Allowing layoffs to follow on-the-job search will make workers initially search more intensively on the job because unsuccessful search exposes them to the subsequent risk of being laid off; thus, by inducing quits and thereby decreasing the number of subsidized workers at low-VMP firms, the benefits of layoffs should be enhanced.

These incentive and screening effects are unique to contracts which prescribe on-the-job search before layoffs because, in this situation, search effort during the first subperiod will depend on the retainment rate during the second (see Figure 2(b)). The derivative of the Lagrangian corresponding to (13), $L^{QL} = W^{QL} + \lambda Z^{QL}$, with respect to the retainment rate is

(14b)

$$\begin{aligned}\frac{\partial L^{QL}}{\partial r} &= (1-t)F_L^1\left[U_L^2 - V^2 + \lambda\left(\pi_L^2 + s\right)\right]\\ &\quad + (1-t)(1-r)F_L^1\left(\frac{\partial V^2}{\partial r}\right)\\ &\quad + \lambda\left\{t\frac{\partial \pi_L^1}{\partial r} + (1-t)\right.\\ &\qquad \left.\times\left[r\pi_L^2 - (1-r)s\right]\left(\frac{\partial F_L^1}{\partial r}\right)\right\}.\end{aligned}$$

Conditional on not quitting during the first subperiod, the first term in (14b) represents the same net benefit described earlier in (14a). The second term in (14b) represents the screening effect and the third term represents the incentive effect. An example in the Appendix (Part B) details these effects and confirms their predicted signs.

B. *The Equilibrium Layoff Rate*

In this part and the Appendix we examine some special versions of our general formulation. In all instances, we assume that workers' preferences and search technologies are characterized as follows:

ASSUMPTION 1: *The search disutility function takes on the form $\beta(e,h) = e$; with $\beta_h = 0$ we know, from* (6) *and* (7), *that long-term and spot employment contracts will exhibit production efficiency.*

ASSUMPTION 2: *The income-leisure utility function takes on the special form $\alpha(w,h) = \mu(w - h^2/2)$, where $\mu(0) = 0$. We shall refer to $w - h^2/2$ as the "net income" of a job, that is, the wage income net of labor costs. Since production efficiency implies $-\alpha_h/\alpha_w = \theta$, workers will supply $h_i = \theta_i$ hours when $\theta = \theta_i$. Thus, $\alpha(w_i, h_i) = \mu(w_i - \theta_i^2/2)$ in equilibrium. Observe that the assumed form of the utility function implies $h_H > h_L$ and hence results in work-sharing.*

ASSUMPTION 3: *We consider the following special case of the search process introduced earlier: An employed (laid-off) worker either expends at least e_{on} (e_{off}) to become fully informed of all employment opportunities, or foregoes search and remains uninformed (and immobile). The requisite effort levels, e_{on} and e_{off}, are fixed for each worker but can vary across workers with the same employment status.*

Assumption 3 allows us to endogenize workers' alternative job offers in the *ex post* spot market as follows.

With firm-specific shocks and constant returns, there will always be a group of high-VMP firms that is willing to hire workers on spot contracts. Since all searching workers have perfect information, the resulting *ex post* equilibrium must conform to the classical zero-profit competitive outcome; hence all spot contracts must pay $\bar{z} = \theta_H^2$ for $\bar{h} = \theta_H$ units of labor, so that $\alpha(z, \bar{h}) = \mu(\theta_H^2/2)$.

Thus, using our earlier notation, the on-the-job search technology corresponding to Assumption 3 can be formally described as follows:

$$F(z,e) = -\begin{cases} 1 & z \geq 0, \quad e < e_{\text{on}}, \\ 1 & z \geq \theta_H^2, \quad e \geq e_{\text{on}}, \\ 0 & z < \theta_H^2, \quad e \geq e_{\text{on}}, \end{cases}$$

where $F(z,e)$ is the probability that a worker's best offer is z or less, given search effort e on the job. The off-the-job distribution function, $G(z,e)$, is described similarly with e_{off} replacing e_{on}.

Layoffs Without Quits.[26] In this subsection we assume that layoffs precede quits and search occurs at most once ($t = 1$), that workers are identical (the search cost distributions are degenerate), and that $e_{\text{on}} \geq \mu(\theta_H^2/2)$ while $e_{\text{off}} = \hat{e} < \mu(\theta_H^2/2)$. In this benchmark case, it is clear that retained workers will choose to forego search (and not quit); the Appendix allows nondegenerate cost distributions, so that some retained workers quit, and compares situations where layoffs precede and follow quits.

If laid-off workers chose to forego search, so that $V(s) = \mu(s)$, it is *as if* they are immobile; and, when laid-off workers are immobile, it is easy to show that the optimal layoff rate is zero.[27] Hence layoffs occur only when the level of severance pay is set so that laid-off workers search. Retained and laid-off workers' expected utilities are therefore, respectively, $U(w_i, h_i) = \mu(w_i - \theta_i^2/2)$ ($i = L, H$) and $V(s) = \mu(s + \theta_H^2/2) - \hat{e}$. Since the equilibrium contract maximizes total expected utility,

$$W = r\mu\left(w_L - \theta_L^2/2\right) + (1-r)$$
$$\times\left(\mu\left(s + \theta_H^2/2\right) - \hat{e}\right) + \mu\left(w_H - \theta_H^2/2\right),$$

subject to the zero-expected profit constraint,

$$Z = r\left(\theta_L^2 - w_L\right) - (1-r)s + \left(\theta_H^2 - w_H\right) = 0,$$

we see that employed workers will receive full insurance, that is, $\mu'(w_L - \theta_L^2/2) = \mu'(w_H - \theta_H^2/2)$. If, as we assume here, negative severance payments are not permitted, it can be shown that the equilibrium value of s must be zero.[28] (Negative values are discussed below.)

The model's remaining equilibrium properties are derived as follows: Solving $Z = s = 0$ and $w_L - \theta_L^2/2 = w_H - \theta_H^2/2$ gives

$$w_i - \theta_i^2/2 = \left(r\theta_L^2 + \theta_H^2\right)/2(1+r) \equiv \delta(r),$$

so that workers' expected utility W can be rewritten as

$$W(r,\hat{e}) \equiv (1+r)\mu(\delta(r))$$
$$+ (1-r)\left(\mu\left(\theta_H^2/2\right) - \hat{e}\right).$$

Observe that $W_{rr} < 0$ for $0 \leq r \leq 1$. There are

---

[26] A more detailed exposition of the material in this section, including some omitted proofs, is contained in Arnott, Hosios, and Stiglitz, 1983. The model in this section resembles Baily's, 1977.

[27] This is because a contract that pays w for h hours to retained workers, pays s to immobile laid-off workers and uses the retainment rate $r < 1$ is dominated by the contract that pays $rw + (1-r)s$ for rh hours to retained workers and retains all workers, that is, $r(\theta h - w) - (1-r)s = \theta rh - [rw + (1-r)s]$ while $r\alpha(w,h) + (1-r)\alpha(s,0)) < \alpha(rw + (1-r)s, rh)$.

[28] Since $\mu(\theta_H^2/2) - \hat{e} > \mu(0) = 0$, every laid-off worker will search when $s = 0$. To verify that $s = 0$, note that $w_L - \theta_L^2/2 = w_H - \theta_H^2/2$ and $\theta_H^2 > \theta_L^2$ imply $\theta_H^2 - w_H > \theta_L^2 - w_L$, so that $Z = 0$ and $s \geq 0$ give $\theta_H^2 - w_H \geq 0$. Therefore, in equilibrium, the profit per employee at high-VMP firms with this implicit contract is greater than or equal to zero, while the profit per employee at firms with spot contracts is zero. Since the spot contract maximizes $\mu(s + w - h^2/2)$ subject to $\theta_H h - w = 0$, and since the optimal implicit contract maximizes $\mu(w_H - h_H^2/2)$ subject to $\theta_H h_H - w_H = \varepsilon$, for some $\varepsilon \geq 0$, $s > 0$ implies $\mu(s + \theta_H^2/2) > \mu(w_H - \theta_H^2/2)$. Hence a nonpositive s value is required to equate retained and laid-off workers' marginal utilities of income, $\mu'(w_H - \theta_H^2/2)$ and $\mu'(s + \theta_H^2/2)$.

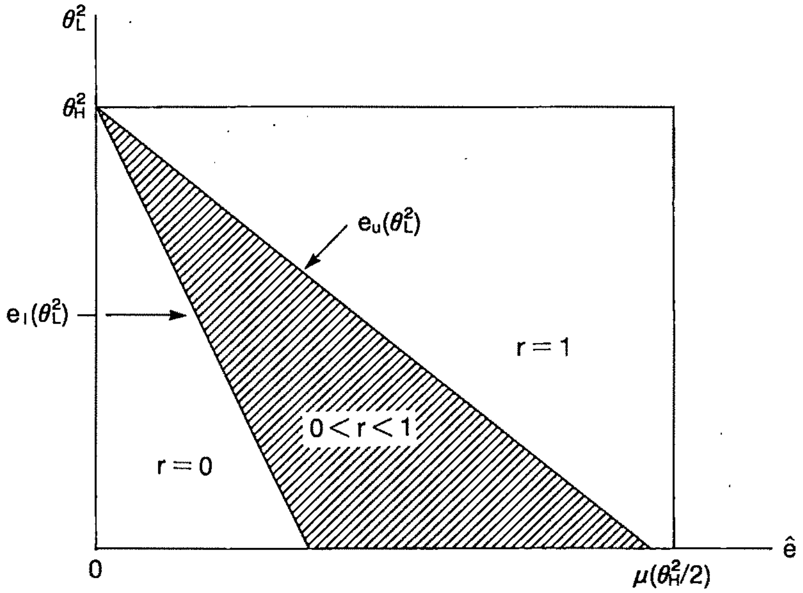


FIGURE 3

two corner solutions to consider: if $W_r \leq 0$ at $r=0$, no one is retained; and if $W_r \geq 0$ at $r=1$, everyone is retained. To determine when these solutions occur, define the critical effort levels, $e_u(\theta_L^2)$ and $e_l(\theta_L^2)$ by the first-order conditions

$$W_r\big(1, e_u(\theta_L^2)\big) = 0 = W_r\big(0, e_l(\theta_L^2)\big).$$

These effort levels are represented in Figure 3. Thus, if $\hat{e} \geq e_u(\theta_L^2)$, $r=1$; if $\hat{e} \leq e_l(\theta_L^2)$, $r=0$; and if $e_l(\theta_L^2) < \hat{e} < e_u(\theta_L^2)$, $0<r<1$ and it can be shown that $dr/d\hat{e} > 0$ and $dr/d\theta_L > 0$.

The tradeoff between risk-sharing considerations and search costs is clear. When off-the-job search becomes less expensive and hence more efficient, or the negative productivity shock becomes larger (as θ_L falls), the balance tips toward layoffs and search, and away from work-sharing; any combination of $\hat{e}$ and θ_L^2 below (above) the shaded area in Figure 3 involves layoffs (work-sharing) and no work-sharing (layoffs).[29] *For intermediate values of $\hat{e}$, however, some workers are retained to work-share in the low-VMP state while the remainder are laid off with zero-severance pay and undertake search.*

On Severance Pay (Again). With constant returns to labor and identical workers, one might have conjectured that the equilibrium employment contract would treat all workers identically *ex post*, and hence that the equilibrium layoff rate must be a corner solution, either zero or one. In fact, of course, interior solutions are possible, as demonstrated above.

Thus far we have restricted severance pay to the nonnegative (due to enforcement problems) and hence one might have conjectured further that the binding nonnegativity constraint in our analysis is the source of the interior solution. It can be shown that this is not in fact the case: *When negative severance pay is allowed, there is full insurance and interior layoff solutions* $(0<r<1)$ *are possible whenever search involves some nonpecu-*

---

[29] These corner solutions are familiar: when workers are freely mobile ($\hat{e}=0$), one-period implicit contract models must replicate spot auctions; and when workers are immobile ($\hat{e} > \mu(\theta_H^2/2)$), the concavity of μ ensures that work-sharing strictly dominates layoffs.

niary and hence uninsurable costs. With complete insurance and *only* pecuniary search costs, however, the equilibrium layoff rate simply implements the standard "first-best" allocation ($r = 0$ or 1).[30]

Layoffs with Quits. The model described above was structured so that there would be no on-the-job search (and hence no quits). This model allowed us to identify a simple inverse relationship between the cost of off-the-job search and the equilibrium layoff rate and, as a special example, confirmed that interior layoff solutions are a feature of the general class of models described earlier in Section II, Part A. Nevertheless, this model failed to capture the tradeoff between voluntary separations (quits induced by lower pay and underemployment) and involuntary separations (through layoffs).

In the Appendix we allow quits and describe the factors that determine the equilibrium mix of quits and layoffs. To capture the sorting effect of quits in the simplest way, we modify the above model by assuming that search costs differ among workers. In particular, we assume that immediately after learning his employment status, each retained worker draws a private value for e_{on} from a distribution with c.d.f. $I(e_{\text{on}})$; and each laid-off worker draws a private value for e_{off} from a distribution with c.d.f. $J(e_{\text{off}})$. (Recall, retained (laid-off) workers must expend $e \geq e_{\text{on}}$ ($\geq e_{\text{off}}$) to find alternative employment.)

Although quits expose workers to less down-side risk and induce voluntary sorting when workers differ in the ability at search, we show that layoffs are doubly beneficial: layoffs force workers to search off the job, which is likely more efficient that on-the-job search; and, when on-the-job search precedes layoffs, they induce more intensive on-the-job search and hence more turnover at low-productivity firms. The former benefit is highlighted in Part A of the Appendix, by comparing the effects of different $I(\ )$ and $J(\ )$ distributions on the layoff rate, using a version of the model in which layoffs precede quits; the latter incentive effect of layoffs and the companion screening effect of on-the-job search are highlighted in Part B with a version of the same model in which quits precede layoffs. Since workers prefer the quit-layoff sequence whose optimal contract prescribes layoffs, however, the analyses in this section and in the Appendix indicate that whether layoffs precede or follow quits, one should nevertheless expect to observe both quits and layoffs at low-productivity firms.

III. Concluding Remarks

A complete theory of unemployment should answer the following questions: How do we explain the degree of observed wage flexibility? How do we explain the form of employment changes, for example, variable hours versus layoffs? How do we explain which workers get laid off? How do we explain the level of severance pay? Why does severance pay not fully compensate laid-off workers? Why do workers prefer to be retained rather than laid off? Why are some unattached workers unable or unwilling to secure employment elsewhere?

While existing implicit contract models answer some of these questions, they fail to answer others and, in some instances, yield counterfactual implications. We have argued that costly search coupled with firms' inability to monitor workers' search activities can provide insights into each of the questions listed above; they can explain partial wage insurance, underemployment at low-VMP firms, layoffs, and quits, sluggish interfirm mobility, and equilibrium unemployment.

A complete theory of unemployment should also indicate whether the allocation of resources is efficient. From the literature on equilibrium with incomplete markets and imperfect information, it should be evident that although the equilibrium contracts described in this paper are "locally efficient" (i.e., given the actions of all other firms in the economy, these contracts maximize workers' expected utility subject to a zero-profit constraint), the market equilibrium, and the corresponding "natural rate" are generally not constrained Pareto-efficient.[31]

---

[30] That is, no (all) workers are laid off at low-VMP firms if the return from staying is greater (less) than the return from alternative employment net of search costs, that is, if $\theta_L^2/2 \geq (<) \theta_H^2/2 - \hat{e}$, where $\theta_i^2/2$ is a worker's net income from employment ($w - h^2/2$) when his MRS between wages and hours equals the MRT ($h = \theta_i$) and he is paid his marginal product ($w = \theta_i$, $h = \theta_i^2$).

To summarize: By highlighting the implications of private search information, this paper has resolved several of the outstanding conundrums in the implicit contract literature and has also provided a unified treatment of variable hours, layoffs, severance pay, quits, and unemployment. The reason why it is necessary to formulate a model which incorporates all of these features should be apparent; without doing so, one cannot be sure that whatever results one obtains are not due to artificial restrictions (for example, that there is no severance pay, no on-the-job search, or that layoffs precede search on the job).

Our model can be viewed as an extension of the standard theory to take into account that what is at risk is not being laid off, per se, but rather being laid off and not rehired. As an extension of this earlier theory, however, some common issues remain; how, for instance, are such contracts enforced? While these problems remain important, their resolution will not, we believe, alter the qualitative insights provided by our analysis.

APPENDIX

In this Appendix we examine different versions of the model described in Section II, Part B; to simplify the remaining analysis we ignore severance pay.

A. *Layoffs Precede Quits.* We assume that immediately after learning his employment status, each worker draws a private value for e_{on} if he is retained or e_{off} if he is laid off. We suppose that retained workers' e_{on} values are drawn from a distribution with continuous c.d.f. $I(e_{\text{on}})$ and support $[0, \mu(\theta_H^2/2)]$ and that unattached workers' e_{off} values are drawn from another distribution with the same support and a possibly different mean value $\hat{e}$.

Assuming once again that retained workers search to quit (rather than quit to search) and substituting the equilibrium labor supplies $h_i = \theta_i$, agents' expected utilities are again described by $W(C)$ and $Z(C)$ as in Section I, except that

$$U(w_i, h_i) = (1 - I(\chi_i))\mu\left(w_i - \theta_i^2/2\right)$$

$$+ I(\chi_i)\mu\left(\theta_H^2/2\right) - \int_0^{\chi_i} e\,dI(e),$$

$$V(0) = \mu\left(\theta_H^2/2\right) - \hat{e},$$

$$\pi(w_i, h_i, \theta_i) = (1 - I(\chi_i))\left(\theta_i^2 - w_i\right),$$

where $\chi_i \equiv \mu(\theta_H^2/2) - \mu(w_i - \theta_i^2/2) > 0$. χ_i is the difference in utility from employment in a new job compared to that in the current job; hence, on-the-job search is profitable if and only if $e_{\text{on}} < \chi_i$. Thus, retained workers quit with probability $I(\chi_i)$ while all laid-off workers seek and find employment.[32]

As we have emphasized, there are two different ways for workers to leave low-VMP firms, by quits or layoffs. The former has the advantage of selecting workers with the lowest e_{on} values for search; the latter has the potential advantage of forcing workers to use a relatively more efficient off-the-job search technology and of generating mobility even with close to full insurance. To illustrate these effects and the possibility of an interior layoff solution, we evaluate the derivative of the

---

[31] The term *constrained* Pareto-efficient is used to indicate that the planner confronts the same informational problems as do private agents.

Arnott and Stiglitz, 1985, have shown that implicit contracts with job turnover are not constrained Pareto-efficient because of the externality exerted by firms' wage/employment decisions, via their effect on workers' savings behavior, on other firms' profits. Two related problems should also be mentioned: First, although we have not formally modeled the micro-structure of search, it is clear from the work of Peter Diamond, 1982, and Dale Mortensen, 1982, that were we to do so, a variety of search externalities could also arise. Second, with risk-averse individuals and more than one commodity, the market equilibrium with limited risk markets will not in general be constrained Pareto-efficient; building on work by David Newbery and Stiglitz, 1982, Hosios, 1984, describes the welfare economics of employment contracts when risk markets are incomplete.

[32] That is, since $\mu(0) = 0$, it always pays a laid-off worker to search since $\mu(\theta_H^2/2) - e \geq 0$ for all e in $[0, \mu(\theta_H^2/2)]$. Also, note that, in the present model, lowering wage w_L encourages quits by raising the maximum search effort level χ_L consistent with a profitable move; thus, as before, $q_w < 0$ and hence our basic incomplete insurance result, (3), holds here as well.

Lagrangian, $L = W + \lambda Z$, where

$$\frac{\partial L}{\partial r} = U(w_L, h_L) - V(0) + \lambda\pi(w_L, h_L, \theta_L)$$
$$= -\left[(1 - I(\chi_L))\chi_L + \int_0^{\chi_L} e\,dI(e)\right]$$
$$+ \hat{e} + \left[\lambda(1 - I(\chi_L))(\theta_L^2 - w_L)\right].$$

This is a special case of (14a).

If, at one extreme, off-the-job search is costless ($\hat{e} = 0$), $\partial L/\partial r$ is strictly negative because the first and last terms are strictly negative (since $w_L > \theta_L^2$) while the middle term is zero; thus all workers are laid off. If, at the other extreme, on-the-job search is costless so that $I(\chi_L) = 1$, $\partial L/\partial r$ is strictly positive because the middle term is positive while the first and last terms are zero; in this case all workers are retained.

We conclude that preferences and search technologies can always be found such that the equilibrium layoff rate is positive at low-productivity firms. If, in addition, $d^2L/d^2r < 0$, the equilibrium layoff rate will generally be less than one; that is, some workers are retained and so quits accompany layoffs. (L will typically be concave in r because of diminishing marginal utility of income or (while outside the present model) diminishing returns to labor).

B. *Quits Precede Layoffs*. We now illustrate the screening and incentive effects that occur only when quits precede layoffs. The main features of the model in Part A are unchanged except that the *ex post* period is divided into two subperiods of equal length ($t = 1/2$). At the beginning of the first subperiod, after VMP values are realized, all workers are retained, their e_{on} values are realized, and they search on the job; some of these workers quit to work elsewhere. At the beginning of the second subperiod, a fraction of the remaining workers are laid off and the rest are retained; laid-off workers' e_{off} values are realized and they search off the job. To simplify we assume that on-the-job search takes place only during the initial subperiod.[33]

Under these assumptions, the expected utility of a worker at a low-VMP firm equals

$$(1 - I(\chi_L))\hat{U} + I(\chi_L)\mu(\theta_H^2/2) - \int_0^{\chi_L}(e/2)\,dI(e),$$

where $I(\chi_L)$ is the probability that the worker quits,

$$\chi_L = \mu(\theta_H^2/2) - \hat{U},$$
$$\hat{U} = (1/2)\left[\mu(w_L - \theta_L^2/2) + r\mu(\omega_L - \theta_L^2/2) + (1 - r)V(0)\right],$$
$$V(0) = \mu(\theta_H^2/2) - \hat{e}(\chi_L),$$

ω_L is the wage paid to workers who remain with the firm, and $\hat{e}(\chi_L)$ is the expected off-the-job search cost of workers whose on-the-job cost exceeded χ_L during the initial subperiod. The corresponding expected profit per worker equals

$$(1/2)(1 - I(\chi_L))\left[(\theta_L^2 - w_L) + r(\theta_L^2 - \omega_L)\right].$$

Differentiation of the corresponding Lagrangian L establishes that $w_L = \omega_L$.

Suppose workers prefer to be retained during the second subperiod, that is, $\mu(w_L - \theta_L^2/2) > V(0)$ (see Section I, Part C). Ignoring the common factor 1/2 and differentiating L gives

$$\frac{\partial L}{\partial r} = (1 - I)\left[\mu(w_L - \theta_L^2/2) - V(0) + \lambda(\theta_L^2 - w_L)\right]$$
$$+ \left[(1 - I)(1 - r)\frac{dV}{d\chi_L} + \lambda(1 + r)(\theta_L^2 - w_L) \times \left(-\frac{dI}{d\chi_L}\right)\right]\frac{\partial \chi_L}{\partial r}.$$

This is a special case of (14b). The first term is standard; it equals the expected benefit for workers, $(1 - I)(\mu - V)$, minus the expected cost for firms, $(1 - I)\lambda(w_L - \theta_L^2)$, of a marginal increase in the retainment probability r. The second term is more interesting; it measures the net marginal benefit of increasing the retainment probability that comes about through the induced change in search.

To sign the second term, observe that increasing the on-the-job search cost below which workers quit, χ_L, increases the likelihood that a worker will initially quit; hence $dI/d\chi_L > 0$. Observe also that the initial subperiod of employment screens out the lower on-the-job search cost workers; since it is most likely that workers who did not initially quit because of high on-the-job search costs will also experience high off-the-job costs, we assume that $d\hat{e}/d\chi_L \geq 0$ and therefore that $dV/d\chi_L \leq 0$. Finally, since $\mu > V$, increasing the retirement rate increases the payoff, $r\mu + (1 - r)V$, to *un*successful search on the job; hence increasing the retainment rate will cause a worker who was initially indifferent to forego search, so that $\partial\chi_L/\partial r < 0$.[34] Thus, $\partial V/\partial r \geq 0$ and $\partial I/\partial r < 0$; in this example, therefore, when on-the-job search precedes layoffs, the screening effect of on-the-job search decreases the layoff rate and the incentive effect of layoffs increases the layoff rate.

---

[33] To make models where layoffs precede and follow search comparable, it should be the case that if retained workers can search N times in one situation, then they can do likewise in the other situation. In our simple examples, $N = 1$.

[34] Since $\mu(\theta_H^2/2) - \hat{U}$ is a positive nondecreasing function of χ_L that intersects $\chi_L = \chi_L$ from above, and since raising r decreases $\mu(\theta_H^2/2) - \hat{U}$ for each χ_L value, it follows that the solution of $\chi_L = \mu(\theta_H^2/2) - \hat{U}$ is a decreasing function of r.

REFERENCES

Akerlof, George A. and Miyazaki, Hajime, "The Implicit Contract Theory of Unemployment Meets the Wage Bill Argument," *Review of Economic Studies*, January 1980, *47*, 321–38.

Arnott, Richard J. and Stiglitz, Joseph E., "Labor Turnover, Wage Structures, and Moral Hazard: The Inefficiency of Competitive Markets," *Journal of Labor Economics*, October 1985, *3*, 434–62.

______, **Hosios, Arthur J. and Stiglitz, Joseph E.,** "Implicit Contracts, Labor Mobility, and Unemployment," Discussion Paper No. 543, Queen's University, 1983.

Azariadis, Costas, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, December 1975, *83*, 1183–1202.

______, "Implicit Contracts and Related Topics: A Survey," in Z. Hornstein et al., eds., *The Economics of the Labour Market*, London: HMSO, 1979, 221–48.

______, "Employment with Asymmetric Information," *Quarterly Journal of Economics*, Supplement 1983, *98*, 157–73.

______ **and Stiglitz, Joseph E.,** "Implicit Contracts and Fixed-Price Equlibria," *Quarterly Journal of Economics*, Supplement 1983, *98*, 1–22.

Baily, Martin N., "Wages and Employment under Uncertain Demand," *Review of Economic Studies*, January 1974, *41*, 37–50.

______, "On the Theory of Layoffs and Unemployment," *Econometrica*, July 1977, *45*, 1043–64.

Chan, Ken S. and Ioannides, Yannis M., "Layoff Unemployment, Risk Shifting, and Productivity," *Quarterly Journal of Economics*, May 1982, *97*, 213–29.

Chari, V. V., "Involuntary Unemployment and Implicit Contracts," *Quarterly Journal of Economics*, Supplement 1983, *98*, 107–23.

Cooper, Russell, "A Note on Overemployment/Underemployment in Labor Contracts under Asymmetric Information," *Economic Letters*, 1983, *12*, 81–89.

Diamond, Peter A., "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, October 1982, *90*, 881–95.

______ **and Stiglitz, Joseph E.,** "Increases in Risk and in Risk Aversion," *Journal of Economic Theory*, July 1974, *8*, 337–60.

Geanakoplos, John and Ito, Takatoshi, "On Implicit Contracts and Involuntary Unemployment," Discussion Paper No. 81-155R, University of Minnesota, 1982.

Green, Jerry and Kahn, Charles, "Wage-Employment Contracts," *Quarterly Journal of Economics*, Supplement 1983, *98*, 173–89.

Grossman, Sanford J. and Hart, Oliver D., "Implicit Contracts, Moral Hazard, and Unemployment," *American Economic Review*, May 1981, *71*, 301–07.

Hart, Oliver D., "Optimal Labor Contracts under Asymmetric Information: An Introduction," *Review of Economic Studies*, January 1983, *50*, 3–36.

Hosios, Arthur J., "A Welfare Analysis of Employment Contracts with and without Asymmetric Information," *Review of Economic Studies*, July 1984, *51*, 471–89.

______, "Layoffs, Recruitment, and Interfirm Mobility," *Journal of Labor Economics*, October 1986, *4*, 473–502.

______, "Unemployment and Vacancies with Sectoral Shifts," mimeo., University of Toronto, 1987.

Imai, Haruo, Geanakoplos, John and Ito, Takatoshi, "Incomplete Insurance and Absolute Risk Aversion," *Economic Letters*, 1981, *8*, 107–112.

Ito, Takatoshi, "Labor Contracts with Voluntary Quits," Discussion Paper No. 233, University of Minnesota, 1986.

Kahn, Charles, M., "Optimal Severance Pay with Incomplete Information," *Journal of Political Economy*, June 1985, *93*, 435–51.

Lowenstein, Mark A., "Worker Heterogeneity, Hours Restrictions, and Temporary Layoffs," *Econometrica*, January 1983, *51*, 69–78.

Mortensen, Dale T., "The Matching Process as a Noncooperative Bargaining Game," in John J. McCall, ed., *The Economics of Information and Uncertainty*, Chicago: University of Chicago Press, 1982.

Newbery, David M. G. and Stiglitz, Joseph E., "The Choice of Technique and the Optimality of Market Equilibrium with Rational Expectations," *Journal of Political Economy*, 1982, *90*, 223–46.

Rosen, Sherwin, "Implicit Contracts: A Survey," *Journal of Economic Literature*, September 1985, *23*, 1144–75.

Salop, Steven C., "A Model of the Natural Rate of Unemployment," *American Economic Review*, March 1979, *69*, 117–25.

Stiglitz, Joseph E., "Theories of Wage Rigidity," in James L. Butkiewicz, Kenneth J. Koford, and Jeffrey B. Miller, eds., *Keynes Economic Legacy: Contemporary Economic Theories*, New York: Praeger, 1986, 153–206.

Strand, Jon, "Work Effort and Search Subsidies with Long-Run Equilibrium Contracts," *European Economic Review*, forthcoming.

Weiss, Andrew, "Job Queues and Layoffs in Labor Markets with Flexible Wages," *Journal of Political Economy*, June 1980, *88*, 526–38.

The Stabilization of the U.S. Economy: Evidence from the Stock Market

By MATTHEW D. SHAPIRO*

Is the apparent stabilization of the postwar U.S. economy a "Figment of the Data," as Christina Romer suggests in an important series of papers?[1] The answer to this question has clear and important implications for macroeconomic analysis and policymaking. Major economic time-series as officially published show clear stabilization in the post-World War II era compared to earlier periods. This (apparent) fact is often taken as evidence that economic policy has been effective in reducing the magnitude of economic fluctuations.[2] The policies include activist monetary and fiscal policy and passive built-in stabilizers such as progressive income taxation, unemployment insurance, and deposit insurance.

In studies of a wide range of time-series of interest to macroeconomists and policymakers, Romer finds that each series is too volatile in the early data. The excess volatility occurs for different reasons in the different series; Romer's findings do not point to a single methodological difficulty in the work of early researchers. It is accidental that each series is excessively rather than insufficiently volatile. According to Romer, the main source of the excess volatility in the unemployment series is interpolations of employment and labor force data that do not take into account the procyclicality of productivity and participation. The main source of the excess volatility in the GNP data is an overstatement of co-movements of GNP and output of commodities.

Romer's findings have not gone unquestioned. David Weir (1986) offers an analysis of the original unemployment data that do suggest stabilization since the end of World War II. Nathan Balke and Robert Gordon (1986) present alternative pre-1929 estimates of GNP that also show stabilization. They find that calculations underlying the original GNP data did not overstate the relationship between commodity and noncommodity GNP. Moreover, they offer data on railroad transportation and construction—components of noncommodity GNP—that are more volatile than commodity output.

Further and even more detailed study of the underlying data is perhaps needed to resolve the impasse. I am skeptical, however, that such research would help most economists frame an opinion concerning the debate unless it were to lead to a consensus. The steps in constructing the data are complicated and require judgment.[3] Ultimately, only the participants in the debate may be able to make informed judgments about each of these detailed steps and how they color the final answer. Instead of reexamining the issues raised by Romer (1986b–c), David Weir (1986), and Balke and Gordon (1986), I propose to examine the issue by using a completely different set of data, but one that in theory should be very useful in resolving the question raised by Romer et al. (1986a–c).

I propose to use asset price data to study the volatility of real activity in the U.S. economy after World War II relative to earlier periods. Asset price data are particularly well-suited for addressing this issue for several reasons.

---

*Cowles Foundation for Research in Economics, Yale University, New Haven, CT 06520 and National Bureau of Economic Research. I am indebted to John Campbell, N. Gregory Mankiw, James Poterba, David Romer, Robert Shiller, Lawrence Summers, David Weir, and an anonymous referee for helpful comments and discussion and for financial support from the National Science Foundation under grant no. SES-8521791. This paper was written while I was a NBER-Olin Fellow.

[1] Romer (1986a–c).

[2] Martin Bailey (1978).

[3] For example, some of the difference between Romer, (1986a–c); and Balke and Gordon, 1986, arises through the choice of years for benchmarks.

First, asset price data are usually much better measured than data on real economic activity. Aside from issues of data alignment and construction of index numbers, there is little doubt about what a time-series of asset market data is measuring. There is no reliance on survey or census data, there is no need to interpolate data because prices are observed virtually continuously, and there is relatively little scope for judgment in the compilation of asset market statistics.

Second, data on stock prices are readily available over a long span on a consistent basis. There is no need to splice together series constructed with different techniques from different types of data.

Third, economic theory predicts that there should be a strong link between economic activity and asset values. The stock price is the discounted present value of the firm's payout. Insofar as this payout must ultimately be a function of real activity, there is a link between real activity and stock prices.

Finally, studies of the data confirm that stock prices are related to real economic activity. Stock prices are an important component in the U.S. Index of Leading Indicators.[4] The strong link between stock prices and activity has also been emphasized recently by Stanley Fischer and Robert Merton (1984). R. Officer (1973) finds a significant, positive relationship between estimated volatility of aggregate stock prices and industrial production in post-World War I data.[5]

If the variance of the fundamentals is unchanged, so should be the variance of stock returns. Comparing these variances seems an important step in answering the question raised by Romer et al. (1986a–c).

I. Link Between the Stock Prices and Economic Activity

This paper exploits the link between economic activity and the stock market to learn about the volatility of the economic activity. In this section, I discuss the theoretical linkage between stock price changes and changes in the fundamentals and evaluate some of the pitfalls in making inferences about the distribution of the fundamentals from the distribution of asset market data.

In a simple model, Robert Lucas (1978) shows that there is a unique relationship between the level of economic activity and the value of the stock market. The tight relationship between real activity and asset prices carries over to more complicated and realistic models where no simple solutions are available. Yet, the lessons from this paper do not rely upon a specific model of asset pricing. The null hypothesis is that the distribution of the fundamentals driving the economy has not changed. The validity of evaluating that hypothesis by examining asset returns only requires that the same model apply for all time periods of interest. The power of this approach derives from the quality of the asset market data rather than recourse to a parametric model of the relationship between fundamentals and asset prices.

In the conventional valuation model, stock prices are the present discounted value of future dividends. Despite doubts raised about how well this model fits the data (Robert Shiller, 1981; and N. Gregory Mankiw, David Romer, and Matthew Shapiro, 1985), as long as stock values bear some relation to fundamental values, examination of stock market data should be useful in addressing the question of the volatility of the fundamentals. No rejections of the conventional valuation model suggest the complete absence of a relationship between real variables and asset prices. Indeed, Sanford Grossman and Shiller (1981) find that the consumption-based version of the conventional valuation model is broadly descriptive of U.S. data at least over some subsamples. John Campbell and Shiller (1986) show that there is substantial evidence that the price-dividend ratio reflects the rational expectation of future dividends. Finally, as long as the departures from the valuation model remain the same in the two subsamples, testing whether asset data have equal or greater volatility in the two samples should be useful in testing whether the process for the fundamentals is the same in the subsamples.[6]

---

[4] See *Handbook of Cyclical Indicators*, 1984, p. 68.

[5] Monthly data on industrial production are available beginning in 1919. For this study, one could envision estimating a relationship between output and stock market data in order to translate implications for stock volatility into those for output volatility. Given that doubts about the accuracy of the output data motivate this study, that line of research is not pursued here.

An objection to inferring changes in real activity from changes in market value is that the market value represents a long average of future real activity. Consequently, transitory movements in real activity will have little effect on market value. There is growing evidence, however, that changes in output are dominated by permanent or very persistent components (Charles Nelson and Charles Plosser, 1982; and Campbell and Mankiw, 1987). Under this statistical model of real activity, one would expect a tight relationship between innovations in activity and innovations in market value.

Changes in corporate financial structure will change the variability of stock prices even if the variability of the fundamentals remains unchanged (Fischer Black, 1977). Specifically, an increase in leverage will increase stock volatility. Changes in leverage could therefore cause incorrect inferences about the variability of the fundamentals from the variability of stock returns. Data on the aggregate corporate balance sheet are available in a National Bureau of Economic Research study by Raymond Goldsmith, Robert Lipsey, and Morris Mendelson (1963) for selected years between 1900 and 1958 and in the Federal Reserve Board's Flow of Funds Accounts for years since 1946.[7] Table 1 gives debt-equity ratios for the aggregate U.S. nonfinancial corporate sector for certain years between 1900 and 1985. This ratio is defined as liabilities divided by the difference of assets and liabilities. Assets are valued at replacement cost in current dollars.[8] In the years when the two series overlap, the Federal Reserve Board's series gives a substantially lower ratio.[9] Yet, the series move together closely in years when each is available; earlier data can be scaled down for comparability. Table 1 shows that the debt-equity ratio is substantially lower during the recent period compared with the earlier period. This alone should make the stock market more volatile in the earlier period. Changes in leverage bias tests of the stabilization hypothesis using the stock returns data in favor of the alternative that real activity has become less variable in recent years.

TABLE 1—DEBT-EQUITY RATIO

	Nonfinancial Corporate Sector	
	NBER	FRB
1900	0.753	
1912	1.004	
1922	0.733	
1929	0.731	
1933	0.884	
1939	0.756	
1945	0.542	
1946	0.508	0.321
1950	0.473	0.305
1955	0.496	0.317
1958	0.506	0.328
1960		0.355
1965		0.418
1970		0.464
1975		0.378
1980		0.351
1985		0.468

Sources: NBER: Goldsmith, Lipsey, and Mendelson, 1963, Tables I and Ia, ratio of line III-14 to line IV. FRB: *Balance Sheets for the U.S. Economy*, line 41.

---

[6] All that is required is that variance of the "fad" component remains unchanged. This assumption is appropriate in testing the null hypothesis of no stabilization. If the fad's component has a high variance, the power of the tests will, however, be low.

[7] I have been unable to locate the balance sheet data for years prior to 1900.

[8] This ratio is not the ratio of debt to the market value of equity, which is unavailable in the Goldsmith et al. data. The use of replacement cost valuation of equity to make the leverage comparisons across the decades is problematic only if the mean of Tobin's q differs over the subsamples.

[9] The major difference between the two balance sheets is that the FRB consolidates the corporate sector by netting out most intra-sectoral assets (mainly securities) while the NBER researchers do not net out these assets. See Goldsmith, Lipsey, and Mendelson, 1963, p. 24.

II. Variables and Data

In this section, I discuss the financial variables whose volatility I will study and the data used to measure them. I also address data problems that might arise in using these series to assess changes in volatility. The basic measure of volatility of the stock market will be the standard deviation of return on holding the stock market. The return—defined as capital gain plus dividend yield—is essentially distributed independently over time. Hence, the sample variance should be an unbiased estimate of the population variance.[10] The raw stock market return is a nominal variable. Several approaches are taken to obtaining a real measure. The real return can be measured by deflating dividends and the stock prices by appropriate price deflators. Although this produces a theoretically valid measure of the real return, it does introduce nonfinancial data into the calculation. The excess return on the stock market over a short nominal interest rate can be measured by financial market data alone. Finally, the nominal stock return is of independent interest. The *mean* of the rate of inflation has clearly changed over the last century.[11] If the variance of inflation is unchanged over the two periods, one can compare the variance of the nominal stock returns over the two periods as long as they are computed about different means. In any case, the excess, real, and nominal stock returns yield very similar results.

I also consider the price-dividend ratio. The price-dividend ratio is an attractive measure of movements in the stock market for two reasons. First, the price-dividend ratio can be stationary even if neither variable is. Hence, the price-dividend ratio is a measure of the level of the stock market that has a well-defined variance. Second, it is intrinsically a real measure.

In addition to these returns variables, I present statistics showing movements in dividends. The dividend is ultimately tied to the profits of the underlying capital, so it is potentially an interesting indicator of activity. But the year-to-year link between the two may be weak if corporate dividend policy makes dividends respond slowly to changes in profits. I examine this issue in the next section. I present results for both log detrended and percent change in dividends because of the difficulty of establishing which procedure is the appropriate transformation to render the series stationary.[12]

In summary, the stock returns—nominal, real, and excess—are the major variables to be examined. These are supplemented by the price-dividend ratio and measures of change in dividends. I now turn to a discussion of the specific data used to measure these variables.

The statistics are computed for the period 1872 through 1987 and for various subperiods. Although the underlying data are measured monthly, I consider variances measured for annual data.[13] The stock price is an index of industrial stocks. The stock price index refers to the month of January. For 1871 through 1925, the data are Cowles's (1939) All Stocks Index (series P-1). Since 1926, they are the Standard & Poor's Composite, which extends Cowles's series. The dividend data give the total dividends during the year (Cowles's series D_a-1 for 1871 through 1925 and the Standard & Poor's series since 1926). The returns are calculated for holding the index from January to January and are expressed as percentages.[14]

---

[10] See Mankiw, Romer, and Shapiro (1985), and Terry Marsh and Merton (1986) for a discussion of how serial correlation in measures of the *level* of stock prices (such as Shiller, 1981) can lead to serious bias in estimates of volatility.

[11] Robert Barsky (1987) argues persuasively that the mean inflation rate was essentially zero before 1914.

[12] Campbell and Shiller (1986) reject the hypothesis that log dividends have a unit root in favor of a trend stationary alternative. They find, however, that the log dividend-price ratio is trend stationary while log prices are not. As they point out, these three findings are inconsistent.

[13] More precise estimates of the variance of stock price growth can be obtained by using higher frequency data, but the dividend data required to compute the returns are only available annually in the early period. Variances of stock price growth computed from the monthly data show the same pattern across time as the annual returns data used in this paper.

Except for the possibilities of errors in the transcription of the raw data underlying these series or arithmetic errors in constructing these series, we can be confident that these stock market data are accurate measures.[15] The monthly Cowles index is the average of the high and low for the month rather than the average of all the days. This procedure does not introduce important bias into the variance of the annual returns measure.[16]

The firms represented in the stock market index do not create all the output of the economy. Consequently, how well the stocks in the index represent the economy as a whole needs to be evaluated. The most dramatic change in the economy over the sample period is the reduction in the importance of agriculture. At the beginning of the sample, over half of output was in the agricultural sector; at the end, it was about 2 percent.[17] Clearly, farms are not traded on the stock exchange although the returns to industrial companies are likely to be affected by agricultural shocks. That stock market data exclude the direct contribution of the farm sector is probably a virtue. The debate over the stabilization of the economy centers on changes in government policies and institutions that are unlikely to affect year-to-year movements in agricultural output. Hence, it is appropriate to use measures that apply to the industrial sector as evidence in the stabilization debate.

Within the industrial sector and within the stock market index, the composition of firms has changed over the sample. For example, railroads were much more important in the earlier period than they have been in the recent period. Changes in the mix of stocks in the index should reflect changes in the mix of economic activity. To make inferences about the volatility of the fundamentals from the volatility of asset returns requires that that relationship between the returns on the major companies in the economy and the underlying economic activity has not changed over the sample period. This maintained hypothesis seems an appropriate component of the complicated, composite hypothesis that the volatility of the economy is unchanged. Nonetheless, the limited coverage of the stock price index and of the stock market itself is a drawback of the approach in this paper.

To obtain a real return, stock prices are deflated by an index of the monthly wholesale price and dividends by an index for the calendar year. For 1871–1889, the price index is the George Warren and Frank Pearson (1935, Table 1) series.[18] Since 1890, it is the Bureau of Labor Statistics Wholesale Price Index (now called the Producer Price Index).[19] The stock price is deflated by the January price index; the dividend by the calendar year index. Deflating by this price

---

[14] The return is $100((P_{t+1}+D_t)/P_t-1)$, where P_t is the January price and D_t is the calendar-year dividend. The price-dividend ratio is defined as P_t/D_{t-1}.

[15] Jack Wilson and Charles Jones (1987) examine the monthly Cowles price index for coding errors. They appear to have found an error in the June 1884 price. They also find problems in the Cowles cumulative returns series. They do not find problems with the data used in this paper.

[16] To check this, I carried out a simple simulation. I generated data assuming stock prices followed a geometric random walk at daily frequency with an innovation standard deviation of 16.0 percent annual rate. I generated data for 40 years of data with 300 days per year. I then computed the standard deviation of the "January" to "January" returns using the Cowles and Standard & Poor's procedures for obtaining monthly data. For Cowles, the January value was taken to be the average of the minimum and maximum observation during the first 25 days of the year; for the Standard & Poor's, January was taken to be the average of the first 25 days. In 1000 replications, the mean absolute deviation of the standard deviation of the return measured by either the Cowles or Standard & Poor's procedure averaged about 3 percent of the true standard deviation.

[17] In 1869, agricultural output was 2.54 billion dollars and total output was 4.83 (*Historical Statistics*, series F238 and F240). In 1986, agricultural output was 93.0 billion dollars and the total was 4,235.0 billion dollars (*Economic Report of the President* 1988, Table B-10).

[18] The annual Warren and Pearson series is also in *Historical Statistics of the United States*, Series E-52.

[19] The annual wholesale price index is Series E-23 in *Historical Statistics*. The monthly series is in the BLS LABSTAT file beginning in 1913. The monthly series from 1890 through 1912 is from the Bureau of Labor Statistics, *Index Number of Wholesale Prices on Pre-War Base, 1890–1927* (Washington: USGPO, 1928).

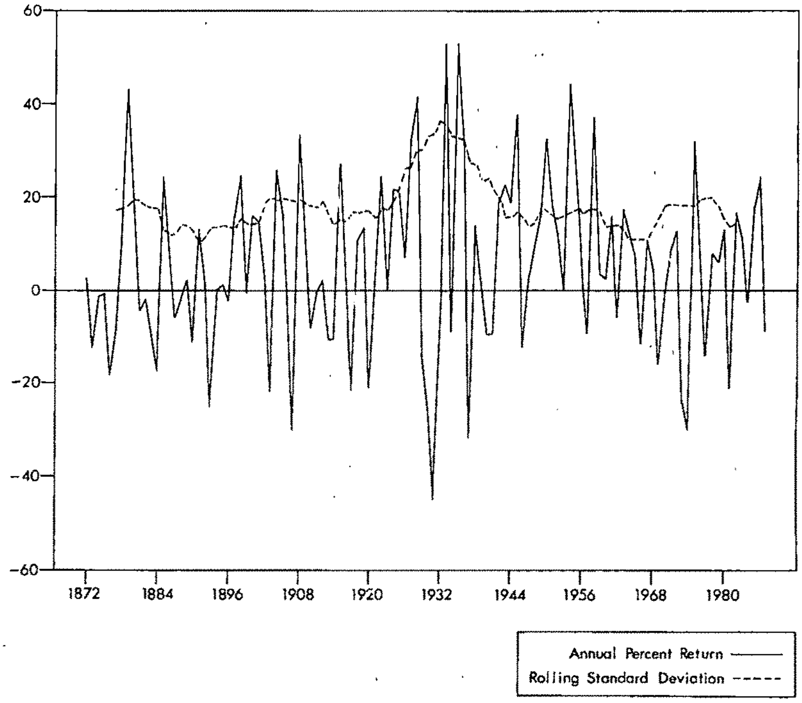


FIGURE 1. EXCESS STOCK RETURN

index might introduce a number of problems in calculation of the returns volatility. Measured price indexes might be too stable if they are based on posted or administered prices rather than market prices. On the other hand, the producer price index might be an excessively volatile measure of inflation. First, producer prices are more variable than consumer prices, at least in recent data.[20] Second, there is an important bias in the retrospective work to construct the historical data. The price data that are easily collected tend to be on commodities that are traded in asset markets. These commodity prices are certainly more volatile than average producer prices. Therefore, it is very difficult even to sign the bias in the volatility of the inflation rate. In practice, the volatility of the nominal returns swamps the inflation volatility in the calculation of real returns.

Moreover, as noted above, consideration of the excess return avoids having to introduce measurement errors in a price index. For this span of data, the short-term commercial paper rate provides a measure of a short-term, low-risk interest rate. The interest rate is the annual six-month commercial paper rate on a bank discount basis published in the *Federal Reserve Bulletin*. It is linked to the four- to six-month commercial paper rate in Milton Friedman and Anna Schwartz (1982).

III. Results

The basic facts about the variability of the asset market data emerge clearly in Figures 1 through 5, which present the levels and rolling estimates of the standard deviations of the excess and real stock returns, the price-dividend ratio, dividend growth, and detrended dividends. The levels of the series are plotted as solid lines. The standard devi-

---

[20] Over 1948 through 1988, the standard deviation of annual inflation (January to January) is 5.4 percent measured by the Producer Price Index and 3.5 percent by the Consumer Price Index.

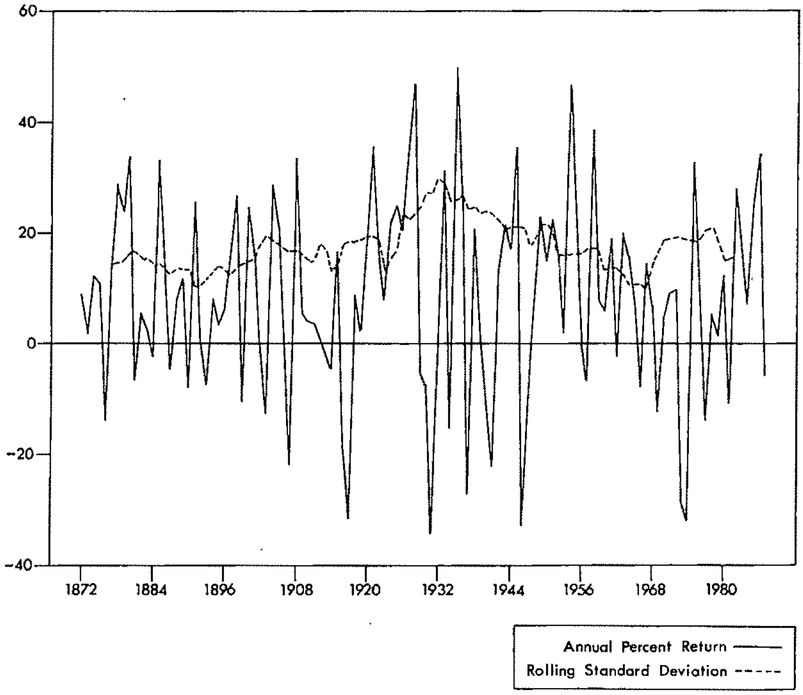


FIGURE 2. REAL STOCK RETURN

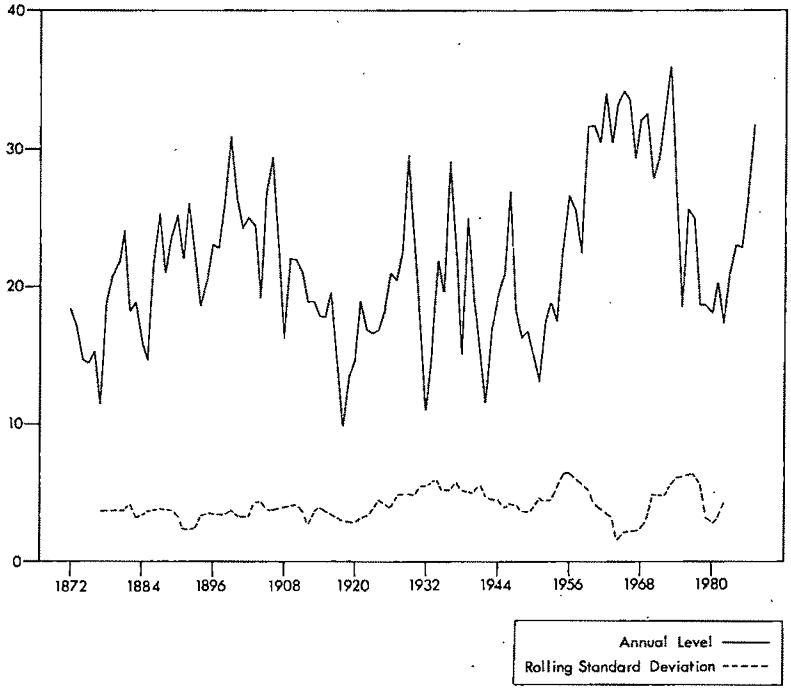


FIGURE 3. PRICE-DIVIDEND RATIO

TABLE 2—STANDARD DEVIATIONS OF STOCK MARKET DATA

	Entire Sample	Subsamples			
	1872–1987	1872–1913	1872–1929	1914–1946	1947–1987
Nominal Stock Return	18.1	15.3	16.3	24.0	15.2
Excess Stock Return	18.7	15.8	16.9	24.2	16.3
Real Stock Return	17.9	14.1	16.0	23.2	16.8
Price-Dividend Ratio	5.8	4.3	4.5	4.7	6.6
Real Dividend Growth	12.4	10.9	12.5	18.2	7.5
Real Dividend Growth Over Four Years[a]	6.2	5.3	6.6	9.2	4.2
Real Log Dividends, Detrended	19.8	14.7	19.3	26.3	17.2

[a] Four-year growth rates expressed as annual rates. Subsamples begin in 1875, 1917, and 1950 to allow for extra lags.

ations of the series are computed over rolling, 11-year sample intervals. These are plotted as broken lines at the midpoint of the intervals. There is no apparent stabilization of either the real or excess stock returns, although volatility increased in the interwar period.[21] The price-dividend ratio is more volatile in recent years.[22] Dividend growth alone does show some stabilization after World War II.

Sample statistics confirm what the eye sees in the figures. Table 2 gives the sample standard deviations of the series for the entire sample and for selected subsamples. The subsamples are 1872–1913, 1872–1929, 1914–1946, and 1947–1987. The first two subsamples are based on alternative ending dates for the earlier period. The first subsample ends with the beginning of World War I, the collapse of the classical gold standard, and the founding of the Federal Reserve. The second subsample ends with the stock market crash. The third subsample includes the turbulent years from the beginning of World War I through the recession following World War II.

The standard deviations of both real stock returns and of the price-dividend ratio are lower in the early period (ending in either 1913 or 1928) than in the period after World War II. Excess returns have about the same variability in the earlier periods and the recent period. The variability of the stock market is consistent with the hypothesis that real activity in the United States has not been stabilized. Only dividend growth shows substantial stabilization in the recent period.

Table 3 gives formal tests of the null hypothesis that the variance of the series in the 1872–1913 and 1947–1987 subsamples are the same. The first column presents the ratio of the variances in the two sample periods. The second column gives a test statistic for the hypothesis that the variances are equal.[23]

---

[21] Officer also finds that the Depression years were especially volatile. He emphasizes that stabilization of the stock market following the Depression cannot be attributed to the creation of the Securities Exchange Commission because post-World War II volatility is no greater than that in the first two decades of the twentieth century. G. William Schwert (1987) finds that high frequency changes in stock volatility are not closely related to changes in volatility of either macroeconomic variables or leverage. He does find increases in volatility during the interwar period.

[22] Note that changes in the price-dividend ratio presented in Figure 3 can be interpreted in terms of changes in the required rate of return. High price-dividend ratios correspond to low required rates of return. Figure 3 suggests that the required rate of return was lower in the post-World War II period than in the pre-World War I period, but that both the early and later periods had lower required rates of return than the middle period. James Poterba and Lawrence Summers (1986) examine the relationship between volatility, stock returns, and the required rate of return. They find volatility changes are not persistent enough to yield large changes in the required rate of return.

[23] Because the series may be serially correlated (serial correlation is nontrivial in the price-dividend ratio and

TABLE 3—TESTS OF CONSTANT VARIANCE, 1872–1913 VS. 1947–1987

	Variance Ratio[b]	$\chi^2(1)$[c]	Probability[d]
Nominal Stock Return	1.01	< 0.01	0.96
Excess Stock Return	0.94	0.07	0.79
Real Stock Return	0.71	3.16	0.08
Price-Dividend Ratio	0.42	13.89	< 0.01
Real Dividend Growth	2.09	2.29	0.13
Real Dividend Growth Over Four Years[a]	1.62	1.12	0.29
Real Log Dividends, Detrended	0.73	0.58	0.45

[a] See note a, Table 2.
[b] Ratio of variance for 1872–1913 to 1947–1987.
[c] Test statistic for hypothesis that variances are equal (see fn. 23).
[d] One minus the probability of rejecting the null hypothesis that the variances of the respective series are the same in 1872–1913 and 1947–1987.

The third column gives one minus the probability of rejecting that the variances are equal. The test statistics confirm what is apparent from the figures and from Table 2. Only for dividend growth can one come close to rejecting the hypothesis that the variance of the series is the same in favor of the hypothesis that it is smaller in the more recent data.[24]

Of the asset market data, only dividend growth shows stabilization in the post-War period. Dividends paid are ultimately constrained by the real return on the underlying asset. But as Marsh and Merton (1986) stress, the timing of dividend payments is largely at the discretion of management. The smoothing of dividends in the post-War period may well reflect a change in dividend policy rather than a change in the underlying driving process. Suppose that managers, for example, decided to spread over several years changes in dividends that would have previously been made over one year. This change in policy would have little effect on the distribution of stock returns. Yet it would reduce the variance and increase the persistence of year-to-year changes in dividends. It is clear from Figure 4 that the serial correlation of the dividend growth has increased since World War II.[25] This increased serial correlation is consistent with more smoothing of dividend changes. Tables 2 and 3 include statistics for the change in dividends measured over four years (expressed at annual rate). Although there is a decline in the variability of this series, it is much smaller than for the year-to-year changes. As Figure 5 and Tables 2 and 3 show, real log, detrended dividends show no evidence of stabilization.

Finally, all the measures considered show dramatic increases in volatility in the period between the beginning of World War I and the end of World War II. The finding accords with the volatile history of that period (the roaring twenties, the Great Depression, the

---

the dividend series), Goldfeld-Quandt tests of equal variance are not appropriate. The probabilities reported in Table 3 are for a test that is valid under very general forms of serial correlation. It is constructed by regressing squared deviations from subsample means on dummies for the subperiods and testing whether the dummies have equal coefficients. The test statistic is based on a covariance matrix corrected for heteroskedasticity and autocorrelation (8-quarter lags) within the subsample but assuming that observations are uncorrelated across subsamples. See Donald Andrews and Ray Fair (forthcoming, 1988) for a general treatment of tests for structural change with heterogenous processes.

[24] The test for real dividend growth illustrates the importance of taking into account the serial correlation in the time-series. Even though the variance of real dividend growth is half of its pre-World War I value in the post-World War II period, the difference is not statistically significant using conventionally sized tests.

[25] The first-order serial correlation of dividend growth is essentially zero in the early period and is 0.30 in the later period.

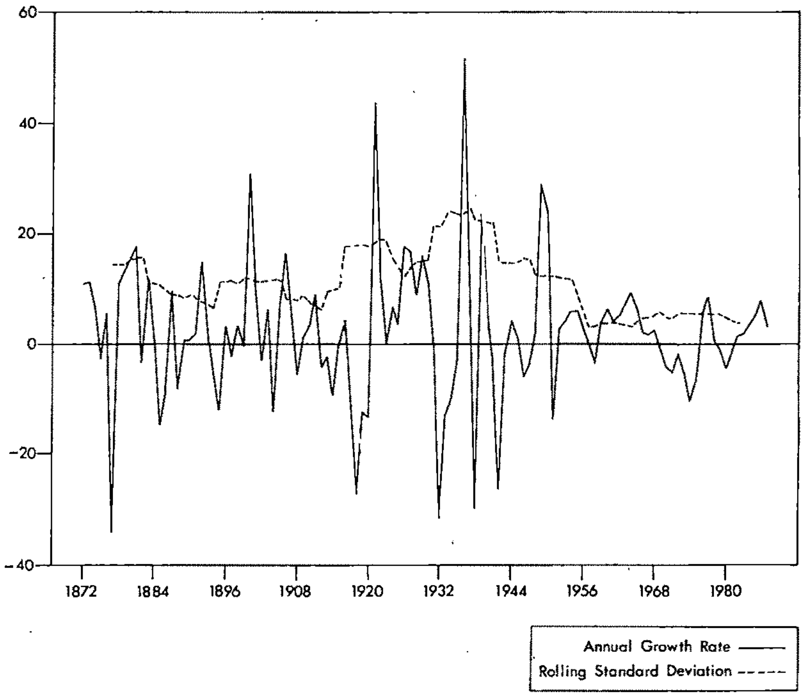


FIGURE 4. REAL DIVIDEND GROWTH

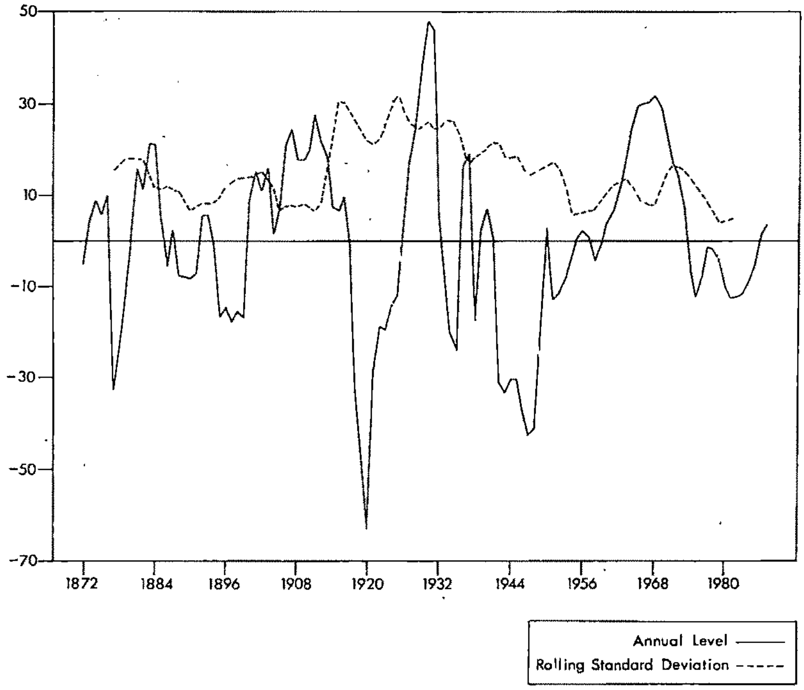


FIGURE 5. REAL LOG DIVIDENDS

World War II boom). The participants in the debates over the historical economic data do not dispute that this period saw an increase in volatility of real activity. Additionally, the large decline in the value of the stock market itself at the onset of the Depression would itself increase volatility through the leverage effect.

IV. Discussion

Section III documents that stock market returns show no reduction in variance when the pre-World War I or pre-Depression periods are compared to the post-World War II period. The lack of stabilization in stock prices supports the view that activity has not been stabilized.

It is inappropriate, of course, to offer evidence of the inability to reject a hypothesis as evidence for the hypothesis. Indeed, if one believes that the stock market is determined by fads or "will-o'-the-wisps" one might maintain that the test lacks power against the alternative that the variance of the fundamentals has changed. The variance of fundamentals could be overwhelmed by the variance of the fad component. In the worst case, all changes in stock market value are caused by speculative bubbles, but there is, as discussed above, evidence that does link the fundamentals to stock prices. Indeed, about 40 to 70 percent of the standard deviation of the log dividend-price ratio can be explained by the rational expectation of future dividends and interest rates in the context of the conventional valuation model.[26] Given that the size of any fad component is thus circumscribed, the tests presented in this paper should have power against the alternative that the variance of the fundamentals has changed.

Another issue of power may arise if the stochastic process for output has shifted during the sample. The greater the persistence of the shocks to the fundamentals, the more the stock market will respond to an innovation of a given size. Campbell and Mankiw (1987) find evidence that the persistence in shocks to output growth has increased in the United States since World War II. Consider the implications of the persistence of output fluctuations for the relationship between variance of output and variance of stock returns. The more persistent are changes in the fundamentals, the greater does the variance of stock returns magnify the variance of innovations in the fundamentals. Conversely, if the fundamentals are negatively autocorrelated, the stock return will respond little to an innovation in the fundamentals. In the limiting case where shocks to the fundamentals are perfectly transitory, for example, where a shock this period is offset by a shock of equal and opposite present value the next period, changes in fundamentals lead to virtually no changes in stock values. Suppose macroeconomic stabilization policies have reduced the variation in output by attenuating high frequency, negatively autocorrelated changes in output.[27] Under this view of policy, stabilization of output would leave the distribution of stock returns relatively unchanged, but year-to-year output changes could be reduced substantially. Hence, if stabilization policy operates only at the short end of the spectrum, stabilization may have been effective, but would go undetected by the tests in this paper. This line of reasoning also suggests that the debate over stabilization should focus more on high frequency movements in output rather than on total variability. Doing so will be difficult given the weakness and relatively low frequency of the existing data. In any case, the findings of this paper rule out stabilization at the long end of the spectrum. Any defense of the stabilization hypothesis must therefore point to high frequency output fluctuations in the early period that have disappeared in recent years.

---

[26] See Campbell and Shiller (1986). Note that this finding is put forward by a strong proponent of fads models (Shiller, 1984). Campbell and Shiller (1988) do find that a lower fraction of variance in the dividend-price ratio is explained when a long moving average of earnings is included in the information set.

[27] Dividend growth clearly shows greater persistence and lower innovation variance in recent years. As noted above, this change in the stochastic process for dividends could be caused by a change in firms' dividend policies.

In summary, financial data provide an excellent vehicle for testing whether real activity has been stabilized when the direct data on real activity are suspect. The financial data are available over a long span, are accurately measured, and are related in theory and in practice to real activity. Stock returns since World War II have essentially the same variance as in earlier periods. Given the evidence that innovations in output are essentially permanent, the constant variance of stock returns supports Romer's important finding that the stabilization of the post-World War II economy is illusory.

REFERENCES

Andrews, Donald W. C. and Fair, Ray C., "Inference in Nonlinear Econometric Models with Structural Changes," *Review of Economic Studies*, forthcoming 1988.

Bailey, Martin N., "Stabilization Policy and Private Economic Behavior," *Brookings Papers on Economic Activity*, 1:1978, 11–50.

Balke, Nathan S. and Gordon, Robert J., "The Estimation of Prewar GNP Volatility," NBER Working Paper No. 1999, 1986.

Barsky, Robert B., "The Fisher Hypothesis and the Forecastability and Persistence of Inflation," *Journal of Monetary Economics*, January 1987, *19*, 3–24.

Black, Fischer, "Studies of Stock Price Volatility Changes," *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economic Statistics Section*, 1977, 177–81.

Campbell, John Y. and Mankiw, N. Gregory, "Are Output Fluctuations Transitory?," *Quarterly Journal of Economics*, November 1987, *102*, 857–80.

______ **and Shiller, Robert J.**, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," NBER Working Paper No. 2100, 1986. *Journal of Financial Studies*, forthcoming.

______ **and** ______, "Stock Prices, Earnings, and Expected Dividends," *Journal of Finance*, July 1988, *43*, 661–76.

Cowles, Alfred, and Associates, *Common-Stock Indexes*, 2nd ed., Bloomington, IN: Principia Press, 1939.

Goldsmith, Raymond W., Lipsey, Robert E. and Mendelson, Morris, *Studies in the National Balance Sheet of the United States*, Vol. 2, Princeton: Princeton University Press, 1963.

Grossman, Sanford and Shiller, Robert S., "The Determinants of the Variability of Stock Prices, *American Economic Review Proceedings*, May 1981, *71*, 222–27.

Fischer, Stanley and Merton, Robert C., "Macroeconomics and Finance: The Role of the Stock Market," *Carnegie-Rochester Conference Series on Public Policy*, Autumn 1984, *21*, 57–108.

Friedman, Milton and Schwartz, Anna J., *Monetary Trends*, Chicago: University of Chicago Press, 1982.

Lucas, Robert E., Jr., "Asset Prices in an Exchange Economy," *Econometrica*, November 1978, *46*, 1429–45.

Mankiw, N. Gregory, Romer, David and Shapiro, Matthew D., "An Unbiased Reexamination of Stock Market Volatility," *Journal of Finance*, July 1985, *40*, 677–87.

Marsh, Terry A. and Merton, Robert C., "Dividend Variability and Variance Bounds Tests for the Rationality of Stock Market Prices," *American Economic Review*, June 1986, *76*, 483–503.

Nelson, Charles R. and Plosser, Charles I., "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications," *Journal of Monetary Economics*, September 1982, *10*, 139–62.

Officer, R. R., "The Variability of the Market Factor of the New York Stock Exchange," *Journal of Business*, July 1973, *46*, 434–53.

Poterba, James M. and Summers, Lawrence H., "The Persistence of Volatility and Stock Market Fluctuations," *American Economic Review*, December 1986, *76*, 1142–51.

Romer, Christina, (1986a) "Spurious Volatility in Historical Unemployment Data," *Journal of Political Economy*, February 1986, *94*, 1–37.

______, (1986b) "Is the Stabilization of the Postwar Economy a Figment of the Data?," *American Economic Review*, June 1986, *76*, 314–34.

______, (1986c) "The Prewar Business Cycle

Reconsidered: New Estimates of Gross National Product, 1869–1918," National Bureau of Economic Research Working Paper No. 1969, 1986, *Journal of Political Economy*, forthcoming.

Schwert, G. William, "The Causes of Changing Stock Market Volatility," unpublished paper, University of Rochester, September 1987.

Shiller, Robert J., "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?," *American Economic Review*, June 1981, *71*, 421–36.

______, "Stock Prices and Social Dynamics," *Brookings Papers on Economic Activity*, 2:1984, 457–98.

Warren, George F. and Pearson, Frank A., *Gold and Prices*, New York: Wiley & Sons, 1935.

Weir, David, "The Reliability of Historical Macroeconomic Data for Comparing Cyclical Stability," *Journal of Economic History*, June 1986, *46*, 353–65.

Wilson, Jack W. and Jones, Charles P., "A Comparison of Annual Common Stock Returns: 1871–1925 with 1926–85," *Journal of Business*, April 1987, *60*, 239–58.

Board of Governors of the Federal Reserve System, *Balance Sheet for the U.S. Economy, 1946–1985*, C.9, October 1986.

U.S. Bureau of Economic Analysis, *Handbook of Cyclical Indicators*, 1984.

U.S. Department of Commerce, Bureau of Census, *Historical Statistics of the United States*, Bicentennial Edition, Bureau of the Census, Washington: USGPO, 1975.

Are Perceptions of Inflation Rational? Some Evidence for Sweden

By LARS JONUNG AND DAVID LAIDLER*

I. The Problem

The rational expectations hypothesis, as usually set out, states that when agents form expectations about an economic variable, they make use of all available information, in such a way that the expectation in question is an unbiased estimate of the variable's realized value. Applied to the price level, or equivalently the inflation rate, this hypothesis is of central importance to contemporary macroeconomics.[1] It is a key component of both neoclassical models in the style of Robert E. Lucas (1973), and more "Keynesian" overlapping contracts models, such as those of Stanley Fischer (1977), Edmund Phelps and John Taylor (1977), and Taylor (1979). The hypothesis is nevertheless very hard to test. It must usually be combined with other postulates, for example that of clearing competitive markets, in order to yield empirical predictions; and if these predictions turn out to be wrong, rational expectations can always be rescued by attributing such errors to the influence of those other postulates.

The availability of survey evidence on agents' expectations does, in principle, permit us to conduct direct tests of rational expectations which are free of this particular problem, but such tests create difficulties of their own. Even setting aside the matter of relying on the truthfulness of the answers which respondents give to questionnaires, it is hard to make unambiguous inferences about the rationality of *ex ante* expectations about variables by comparing them with *ex post* outcomes. Even evidence of serial correlation in expectational errors, at one time thought to be inconsistent with rational expectations, does not definitely refute the hypothesis, as consideration of the by now widely understood "Peso problem" makes clear.[2]

Suppose we had been forming expectations about the future value of the Mexican Peso against the United States Dollar over a period when there was a positive probability of the Peso's being devalued. *Ex ante* rational expectations of its value would have taken account of this probability. Suppose, however, that devaluation did not occur. *Ex post*, our expectations would appear to be subject to systematic error for as long as the positive *ex ante* probability of devaluation persisted without the event actually taking place. The error in question, though, would not reflect the neglect of any available information in the generation of expectations. The "Peso problem" is not a mere curiosity. It is a particular case of a rather general phenomenon. Whenever there exists a positive probability of a unique event occurring, and whenever that positive probability persists over time without the event in question taking place, then *ex ante* rational expectations about variables, whose values the occurrence of that event would affect, will be systematically erroneous *ex post*. *Ex post* systematic errors in expectations, therefore, do not always amount to conclusive evidence of an *ex ante* lack of rationality about those expectations.

---

*Professor of Economics, Stockholm School of Economics and Professor of Economics, University of Western Ontario, Social Science Center, London, Canada N6A 5C2, respectively. We would like to thank Erling Pettersson for excellent computational assistance. We have benefited significantly from comments by Bengt Assarsson, Roy Batchelor, Axel Leijonhufvud, and Allan Meltzer and seminar participants of the Institute of International Economics, Stockholm and Flinders University, as well as two anonymous referees.

[1] Rational agents must be assumed to know the value of the price level P at any time in the past, for example, at $t-1$. If they form an estimate P^e of the current price level, then this immediately implies an estimate of the recent inflation rate $P_t^e - P_{t-1}$ and vice versa.

[2] On this argument, see Milton Friedman and Anna Schwartz (1982, pp. 556–57).

The problem is pervasive, but it is not universal. Certain variables, notably the general price level, are difficult to observe contemporaneously. Indeed, in Lucas' (1973) model of the aggregate supply curve, the possibility of output varying in the short run as a function of the price level depends crucially on the occurrence of discrepancies between the current value of the price level and agents' rationally formed "expectations" —*contemporaneous perceptions* would be a better phrase—of that value. Such contemporaneous perceptions are not susceptible to "Peso problem" effects. Even if the occurrence or not of some unique event is relevant to the current value of the price level, information about whether or not that event has actually occurred will be available to agents by the time the price level is realized; and, if they are rational, agents will have built it into their *contemporaneous perceptions* of the price level's value even though it could not have been incorporated into their *ex ante expectations*.

Thus, survey data on agents' perceptions of the current price level (or equivalently of the amount by which the price level has changed from some known value at a well-defined date in the past), gathered at a time when official measures of that variable are not available to them, can be used with greater confidence to test the rational expectations hypothesis than can survey data on *ex ante* expectations. Such data are available. In this paper we shall use them to test the hypothesis that agents' perceptions of the current price level, formed at the very time at which the information used to construct official measures of that same variable are collected, but before they are made public, are rational.[3]

II. The Data

The data to which we refer above come from a quarterly survey which has been carried out regularly by the National Institute of Economic Research in Stockholm since July 1979. A unique feature of this survey is that it explores *both* the inflation perceptions *and* inflation expectations of a large representative sample of the Swedish public. About 6,500 people are usually questioned on each occasion, though the sample size has on one occasion (October 1984) fallen to 1500. Individuals remain in the sample for not more than four successive surveys.

The data on perceptions of inflation used in the following tests are generated in the following way. Respondents are first asked about the behavior of "prices in general" during the past twelve months: "Have they risen, fallen, or stayed the same?" Those who answer that prices have fallen or stayed the same, negligible proportions never exceeding, respectively, 0.1 and 2.0 percent of respondents over the period of our sample, when prices did rise continuously, are not questioned further. The vast majority, who have answered "risen" are then asked: "By how much?" They are, that is, asked to give a numerical point estimate of the inflation rate. Respondents are permitted to offer an interval estimate or to answer "do not know" to this question, if they are unwilling to give a point estimate; on average, about 53 percent avail themselves of these options. However, with an initial sample size usually in excess of 6,000, each survey nevertheless yields a large sample of quantitative point estimates of the inflation rate. (October 1984 is again an exception here.) Our measure of perceived inflation is the mean value of these point estimates for each of the 23 surveys taken over the period July 1979–January 1985. Those agents answering that prices have fallen, or stayed the same, and those not offering a quantitative point estimate of inflation are ignored, rather than having a

---

[3] As David Laidler and Michael Parkin, 1975, noted, theoretical and empirical studies on inflationary expectations assume, usually tactily, that economic agents hold uniform expectations with complete certainty. However, empirical work shows that (a) inflationary perceptions and expectations display considerable distribution across respondents, see Lars Jonung, 1981, and (b) they are not held with complete certainty; see, for example, Jonung, 1986. We abstract from these issues here.

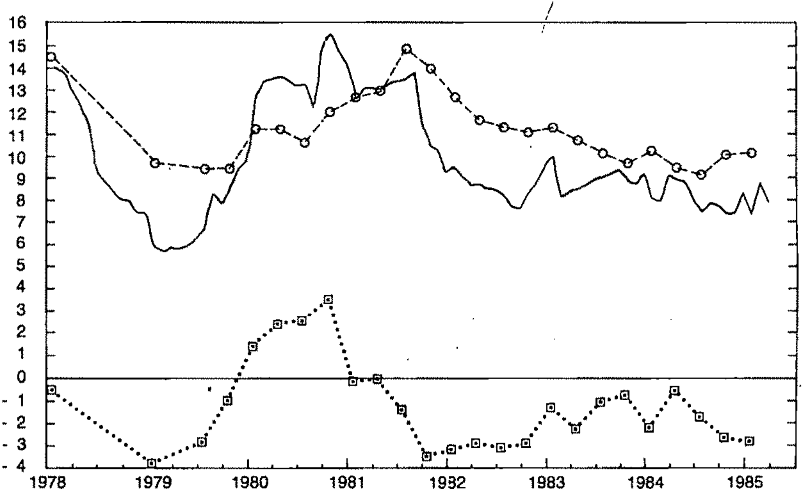


Chart I. Perceived (—∘—) and Actual (—) Inflation Rates, and Their Difference (...□...), January 1978–January 1985

zero value assigned to them.[4] Even so, 54,917 individual answers form the basis of the data we use.

It should be noted explicitly that information about the officially registered consumer price index, (CPI), and thus about the actual rate of inflation during the twelve months preceding each survey, is not available to respondents at the time of the survey. The data on which official price level statistics for the survey month are based are in fact gathered at roughly the same time as the survey is conducted, namely the first two weeks of January, April, July, and October, but the official statistics themselves are not published until a few weeks after the questionnaire has been completed. Our tests are based on comparisons of the actual rate of consumer price inflation as measured by the official CPI and the perceived rate of inflation as measured by the surveys over the period 1978–84. The relevant data are presented in Chart I, where inflation rates, both actual and perceived, are measured as annual percentage changes over the preceding twelve months. The perception error, also plotted in Chart I, is measured as the percentage point difference between the actual and perceived rates of inflation.[5]

---

[4] We believe that, in confining ourselves to data generated by agents who are sufficiently confident in the quality of their own perceptions to offer a quantitative estimate of the inflation rate, we are probably biasing our tests in favor of the rational expectations hypothesis. Note that to drop those respondents not offering a point estimate for the perceived inflation rate from our sample amounts to attributing to them an estimate of inflation equal to the mean of those offering a point estimate. In the case of respondents who perceive prices to have risen, but decline to offer a precise estimate of how much, this procedure seems appropriate. However, it introduces an upward bias to our measure of the perceived rate to the extent that agents who claim prices to have remained the same or fallen are also treated in this way, but the number of such agents is, in fact, negligible. As far as we are aware, quantitative measures of inflationary perceptions have only been used in two previous studies: Jonung, 1981, and Roy Batchelor, 1984. Jonung examines in detail the first survey made for Sweden in 1978. Batchelor transforms qualitative perception data taken from the EEC consumer tendency surveys into quantitative measures of perceptions. His conversion makes actual and perceived inflation rates equal on average, ruling out any possible bias. Even so, he finds that the perceived inflation rate is less volatile than the actual rate—a finding consistent with ours for Sweden. Batchelor interprets this pattern as showing that consumers perceive only a moving average of actual inflation.

[5] The dates of the surveys are marked by circles in the chart. The regular surveys started in July 1979. The

It is immediately apparent from inspection of Chart I that the perceived inflation rate follows a smoother time path than does the actual rate. The latter is characterized by pronounced cyclical movement, and also reflects the influence of various easily identified policy measures, such as the two devaluations of the Swedish currency which took place in August 1981 and October 1982, respectively, as well as a number of changes in subsidies and indirect taxes. In comparison with the actual rate, the perceived inflation rate fluctuates with a smaller amplitude, so that perceptions of inflation are systematically higher than actual inflation when the latter is falling, and systematically lower during periods of rising inflation. The mean perception error for all 25 surveys covered by Chart I is 1.08 percentage points. Evidently, over the sample period, the Swedish public on average overestimated the actual inflation rate, but as we shall see, this small mean prediction error is not statistically different from zero and it would be wrong to attach any economic significance to it. Only a larger set of data, which will become available with the passage of time, can throw further light on whether or not there does exist a systematic long-run bias in the Swedish public's perceptions of inflation.

III. The Tests

The rational expectations hypothesis, as it is usually formulated, implies that the difference between the actual inflation rate and its perceived value is a serially uncorrelated variable with a zero mean. Where p_t is the actual rate of inflation over the 12-month period ending at time t, per_t is the perceived rate of inflation for the same period formed at time t, and u_t is a random error, this version of the hypothesis may be tested partially by estimating the following equation

$$(1) \qquad p_t = a + b\,\mathrm{per}_t + u_t .$$

The economic hypothesis that the perceived rate of inflation over the previous year is an unbiased predictor of the actual—but unknown at time t—rate implies the joint statistical hypothesis that $(a,b) = (0,1)$. This test of unbiasedness is generally considered to be a "weak" test of rationality, because it says nothing about the properties of the error term u_t, which ought, according to the rational expectations hypothesis, to be uncorrelated with any information, including any of its own lagged values, available at the time t to which agents' perceptions refer.

Note that equation (1) is not subject to those common econometric problems which arise when expressions of a similar form are used to test for the unbiasedness and efficiency of *expectations*, and the forecast horizon embedded in those expectations is longer than the interval of observation: for example, when data from quarterly surveys of expectations about inflation over the succeeding twelve months are used.[6] In such a case, an *ex ante* random error in forecasting the inflation rate in a particular quarter can enter into no fewer than four successive forecasts of the annual inflation rate, and hence can induce serial correlation into the errors of the latter. Equation (1), however, deals with *backcasts* rather then *forecasts*. At any time, the actual inflation rate over the first three of the preceding four quarters is public knowledge, available for agents to use in constructing their estimate of inflation over the previous year. Their only scope for error, therefore, arises in misestimating inflation over the immediately preceding quarter.

It is of course possible that agents do not use readily available data regularly to update their backcasts of inflation, and if they fail to do so, this will indeed induce serial correlation into the errors of backcasts of the annual inflation rate made at quarterly intervals. However, given that the data needed for the relevant adjustments are public knowledge, serial correlation induced by such means would be more than a statistical arti-

---

results of two preceding surveys, undertaken in January 1978 and January 1979, are included in the chart but were not used in the regressions reported in Table 1.

[6] On this point see, *inter alia*, Lars Hansen and Robert Hodrick, 1980.

fact, as it is in the analogous forecast case. Rather, it would be evidence that readily available and relevant information was being neglected, and hence that the rational expectations hypothesis, as usually formulated, is erroneous.

Now Chart I clearly indicates that changes in the actual rate of inflation are associated with changes in the perceived rate. It might be thought worthwhile to ask whether the public is able to estimate the rate of change of the inflation rate in a rational way. This question may be answered by fitting an equation of the following form to our data:

$$(2)\quad p_t - p_{t-1} = c + d(\text{per}_t - p_{t-1}) + w_t.$$

Here unbiasedness of expectations implies the joint hypothesis $(c, d) = (0,1)$, while their efficiency again requires that the error term be white noise.

As we shall see in a moment, the outcome of fitting equation (1) is to show that the inflation perceptions of the public are unbiased. However, as inspection of Chart I strongly suggests, the error term of that equation is in fact positively serially correlated. The actual rate of inflation up to the end of the quarter preceding the survey, and the value the inflation rate perceived by respondents up to the end of that previous quarter, represent information freely available to respondents. Hence it is worth asking about whether, and the extent to which, perception errors could be reduced by using information about previous errors. The following relationship between the current perception error and its own lagged value was therefore estimated as

$$(3)\quad p_t - \text{per}_t = e + g(p_{t-1} - \text{per}_{t-1}) + v_t.$$

Here efficiency in the use of information implies the joint hypothesis that $(e, g) = (0,0)$. This test is identical to that used by, among others, Stephen Figlewski and Paul Wachtel (1981) in examining the relationship between recent and past forecast errors of expectations data for the United States.

Of course the lagged value of the perception error is not the only candidate for use as "relevant information" in testing for efficiency of expectations. We also experimented with the following equation

$$(4)\quad p_t - \text{per}_t = e + g(p_{t-1} - \text{per}_{t-1}) + h(x) + v_t,$$

where x represented successively import price inflation (imp) in the previous quarter, and the rate of unemployment (u) in the previous quarter. Efficiency here implies $(e, g, h) = (0,0,0)$. $h \neq 0$ suggests that other information in addition to previous errors could be used to improve the accuracy of inflation perceptions in the current period.

IV. The Results

Table 1 reports the results of fitting equations (1)–(4) to our data. A common property of all these results is that the intercept of the relevant equation is never statistically significantly different from zero. Furthermore, in regression (1), the slope coefficient is not statistically significantly different from one. These parameter values are consistent with inflation perceptions being unbiased. The critical value for F to accept the unbiasedness hypothesis at the 99 percent level of confidence is 5.78, and its actual value is 0.44. However, the low Durbin-Watson statistic for regression (1) indicates the presence of first-order serial correlation in perception errors of the inflation rate. In the presence of such serial correlation, the F test to which we have just alluded is not valid. If we apply the Cochrane-Orcutt adjustment procedure to equation (1) we obtain equation (1′). Here the F test, now valid, yields a value of 0.85 and continues to confirm the unbiasedness of inflation perceptions, but the high value of ρ, 0.763 also confirms the initial impression that perception errors persist over time.

Regression (2) tests hypotheses about changes in the perceived inflation rate. The goodness-of-fit here is low, and more important, the coefficient d is significantly lower than the unit value that unbiasedness in perceptions of changes in the inflation rate would generate. Apparently the Swedish public systematically underestimate changes

TABLE 1[a] —ESTIMATES OF EQUATIONS (1)–(4)

Regression and Dependent Variable	Intercept	Coefficients of Independent Variables		Adjusted R^2	DW
(1) $p_t =$	a	$+b\,\text{per}_t$			
	-1.506	$+\ 1.035$		0.349	0.484
	(3.235)	(0.289)			
(1′) $p_t =$	a	$+b\,\text{per}_t$			
	-1.887	$+\ 1.032$		0.264	1.814
	(3.810)	(0.339)			($\rho = 0.763$)
(2) $p_t - p_{t-1} =$	c	$+d(\text{per}_t - p_{t-1}$			
	-0.319	$-\ 0.309$		0.072	1.537
	(0.403)	(0.191)			
(3) $p_t - \text{per}_t =$	e	$+G(p_{t-1} + \text{per}_{t-1})$			
	-0.275	$+\ 0.752$		0.541	1.698
	(0.332)	(0.148)			
(4) $p_t - \text{per}_t =$	e	$+g(p_{t-1} - \text{per}_{t-1})$	$+h\,\text{imp}_{t-1}$		
	-0.842	$+\ 0.829$	$+\ 0.246$	0.548	1.725
	(0.425)	(0.144)	(0.126)		
(4′) $p_t - \text{per}_t =$	e	$+g(p_{t-1} - \text{per}_{t-1})$	$+hU_{t-1}$		
	1.939	$+\ 0.583$	-0.881	0.573	1.801
	(1.428)	(0.178)	(0.554)		

[a] Figures in parentheses are standard errors.

in the inflation rate. Yet another way of looking at the same problem is provided by regression (3). Here the parameter g may be regarded as an alternative estimate of the parameter ρ obtained from (1′). Its positive value confirms that current errors in perceptions of the inflation rate are significantly positively correlated with their own past values, and could be reduced by paying attention to this readily available information. The critical value for F for accepting the joint hypothesis that $(e, g) = (0,0)$ at the 99 percent level is 5.72, and the actual value of this statistic 18.72, lies well outside this bound. Moreover, when a measure of previous import price inflation is added in equation (4), it turns out to make a marginally statistically significant contribution to improving its goodness of fit. Here again, we have freely available information which could be used to improve inflation perceptions, but which is apparently ignored. The unemployment rate, on the other hand, is a less reliable source of information. However, the unemployment rate is correlated with the lagged inflation perception error and equation (4′) is a slightly better-fitting expression than (4).

To sum up, our empirical results suggest that, unbiased though they are, the Swedish public's perceptions of "actual" inflation, when actual inflation is measured by the CPI, could be rendered more accurate. Freely available and useful information apparently remains unutilized. There are at least two explanations of our results. First, it may simply be that the Swedish price index measures the "true" price level, relevant to the respondents of our questionnaire, imperfectly. If such a hypothetical "true" price index moved more sluggishly than the official index upon which our tests are based, the results which we have generated could be a statistical artifact of no economic significance at all. (However, to minimize this bias the respondents to the survey generating our data were asked to give point estimates of how "prices in general" have moved, not of the behavior of the prices of goods they purchase.) It is impossible to test this suggestion definitively. It would always be feasible to construct *ex post* a price index of whose behavior our "perceptions" data provide an unbiased and efficient estimate. However, we can claim that our results do not depend crucially on the specific price index which we use to measure inflation. In particular, we have rerun our tests using the percentage rate of change of the Swedish food price index to measure inflation, and have found their results to be robust in the face of this change.[7]

Second, and, we would argue, more plausibly, it might be that the costs of gathering and processing information on perception errors, and other data too, are high enough relative to the benefits to be gained from improving inflation perceptions, for it to be *economically* irrational for agents to do this. This interpretation is consistent with recent theoretical work by John Galbraith (1988). Moreover, empirical work now being carried out by Batchelor and Jonung (see 1986 for preliminary results), using data from the same survey underlying our results, shows that differences in the accuracy and purely statistical rationality of inflationary perceptions across gender, age, and income groups appear to be related to the costs and benefits of collecting information on inflation for the various groups examined.[8]

V. Concluding Comments

In this paper we have used survey data on contemporaneous inflation perceptions to test a widely utilized version of the rational expectations hypothesis. Such data enable us to avoid the "Peso problem" alibi for this hypothesis when errors display the serial correlation which is *prima facie* inconsistent with its usual formulation. Even, so, the errors generated in our tests are indeed strongly serially correlated. Though one can never completely rule out the possibility that such results as these have arisen from using the "wrong" data, we are inclined to believe that our tests are more than a mere statistical artifact, and do show that the Swedish public make serially correlated errors when estimating inflation, and do not find it worth the effort to remove them. This interpretation of the evidence is *not* inconsistent with *economic* rationality, because the benefits of improving inflation estimates may not be worth the extra cost of doing so. Nevertheless, our results do suggest that it is inappropriate to rule out a priori the use, in economic modeling, of expectations formulas that lead to serially correlated errors. Though the existence of the Peso problem makes it impossible to test the proposition definitively in the case of *ex ante* forecasts, our results for backcasts at least open up the possibility that persistent expectational errors do, after all, have a role to play, alongside other rigidities, in explaining why the responses of variables such as output and employment to exogenous shocks themselves display persistence in real world economies.

---

[7] Another possible explanation for our results may be that the respondents, when asked about the history of inflation during the past twelve months, actually had in mind when answering the rate of inflation that they perceived to have prevailed over a longer time period. To investigate this view, we calculated the perception error implied by inflation rates measured over varying time periods, starting from 6 months and going back to 36 months from the date of the survey. This perception error turned out to be minimized when we used a 27-month—not a 12-month—measure of the actual inflation rate. This result is consistent with the postulate that respondents were influenced by a history of prices going further back than the immediately prior 12 months referred to in the survey questions. The fact remains, however, that agents were asked about the previous 12 months, not 27 months, and that agents making full use of all available information would have been able to distinguish between the two.

[8] This ongoing work is also addressing a question which naturally presents itself as a followup to the results set out here, namely whether agents' forecasts of inflation display more or less unbiasedness and efficiency than their backcasts. Preliminary results suggest that forecasts, like backcasts, are unbiased, but a comparison of their efficiency is more difficult to make. The forecasts errors made by respondents to our surveys do display serial correlation, but such correlation is in any event induced in those forecasts in a way that it is not in backcasts (see p. 1083 above) by the fact that they are made at quarterly intervals for a 1-year period. Batchelor and Jonung have applied a technique developed by Hansen and Hodrick, 1980, in an attempt to deal with this problem, but in doing so have rendered their results not directly comparable to those presented here, which are generated by applying OLS to backcast data. This exercise yielded results which did not permit the hypotheses of efficiency to be rejected in the case of forecasts, but Batchelor and Jonung attribute this result to a lack of power on the part of the Hansen–Hodrick test when applied to a small sample of data, rather than to any paradoxically proclivity on the part of the subjects of their study to form expectations efficiently and perceptions inefficiently. Clearly however this issue merits further investigation.

REFERENCES

Batchelor, Roy A., "Consumer Perceptions and Expectations," *The Measurement and Impact of Inflation and Growth Expectations*, report to the Commission for the EEC, City University, London, 1984, ch. 5.

______ **and Jonung, Lars,** "Information Theory and Group Differences in Inflation Expectations," seminar paper No. 5, National Institute of Economic Research, Stockholm, November 1986, mimeo.

Figlewski, Stephen and Wachtel, Paul, "The Formation of Inflationary Expectations," *Review of Economics and Statistics*, February 1981, *63*, 1–10.

Fischer, Stanley, "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule," *Journal of Political Economy*, February 1977, *85*, 317–23.

Friedman, Milton and Schwartz, Anna J., *Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates 1867–1975*, Chicago: University of Chicago Press, for the NBER, 1982.

Galbraith, John W., "Modelling Expectations Formation with Measurement Errors," *Economic Journal*, June 1988, *98*, 412–28.

Hansen, Lars P. and Hodrick, Robert J., "Forward Exchange Rates as Optimal Predictors of Future Spot Rates: An Econometric Analysis," *Journal of Political Economy*, October 1980, *88*, 829–53.

Jonung, Lars, "Perceived and Expected Rates of Inflation in Sweden," *American Economic Review*, December 1981, *71*, 961–68.

______, "Uncertainty about Inflationary Perceptions and Expectations," *Journal of Economic Psychology*, September 1986, *7*, 315–25.

Laidler, David and Parkin, Michael, "Inflation: A Survey," *Economic Journal*, December 1975, *85*, 741–809.

Lucas, Robert E., Jr., "Some International Evidence of Output-Inflation Tradeoffs," *American Economic Review*, June 1973, *63*, 326–34.

Phelps, Edmund S. and Taylor, John B., "Stabilizing Powers of Monetary Policy Under Rational Expectations," *Journal of Political Economy*, February 1977, *85*, 163–90.

Taylor, John B., "Staggered Wage Setting in a Macro Model," *American Economic Review Proceedings*, May 1979, *69*, 108–13.

Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation

By HERSCHEL I. GROSSMAN AND JOHN B. VAN HUYCK*

The history of sovereign borrowing suggests the following generalizations about defaults on sovereign debts: (1) Defaults are associated with identifiably bad states of the world. (2) Defaults are usually partial rather than complete. (3) Sovereign states often are able to borrow again soon after default. Barry Eichengreen and Richard Portes (1986) and Jeffrey Sachs (1982) provide recent accounts of this history.

Motivated by these facts, this paper develops a model in which sovereign defaults occur as bad outcomes of debt-servicing obligations that are implicitly contingent on the realized state of the world. The interpretation of sovereign debts as contingent claims implies that sovereign debts serve at least in part to shift to lenders' risks associated with verifiable events that affect the fortunes of the sovereign. This interpretation of sovereign debts also implies that lenders sharply differentiate excusable defaults, which are justifiably associated with implicitly understood contingencies, from debt repudiation, which would be unjustifiable and inexcusable. The extent of the risk shifting provided by sovereign debt depends on the extent of default that is excusable in the event of a bad realization.[1]

A critical aspect of any debt, whatever its risk-shifting component, is the means for validating lenders' servicing expectations—that is, for preventing unjustifiable repudiation. In this regard, private and sovereign debts are fundamentally different. Private debts are subject to laws regarding bankruptcy and enforcement of collateral. Sovereign debts, in contrast, are above the law. Indeed, we can regard the power to abrogate commitments without having to answer to a higher enforcement authority to be the essential property of sovereignty. (Given this definition, historical cases in which one sovereign or its agents seized the assets of another apparent sovereign who is failing to service its debts involve infringement of sovereignty.)

The fact that the servicing of sovereign debts is not subject to external enforcement suggests that the decision by a sovereign not to repudiate its debts depends mainly on the sovereign's concern about its trustworthy reputation for validating lenders' servicing expectations. The analysis that follows formalizes this idea and derives a reputational equilibrium in which the amount of debt and lenders' expectations about contingent debt servicing are such that the sovereign chooses in all states to validate these expectations, thereby maintaining continued access to loans. The derivation of this reputational equilibrium focuses on the contingent debt-servicing policy that lenders can rationally expect the sovereign to follow. This expectation determines whether the reputational equilibrium permits the sovereign to issue an amount of debt sufficient for efficient risk shifting.

---

*Department of Economics, Brown University, Providence, RI 02912 and Department of Economics, Texas A&M University, College Station, TX 77843, respectively. The National Science Foundation under grant no. SES-8408873 has supported Grossman's work. The Social Science Research Council and a Citicorp Graduate Fellowship have supported Van Huyck's work. We are indebted to participants in NBER/FMME Summer Institutes and to anonymous referees for constructive criticisms of earlier versions of this paper.

[1] For economic analysis, default is usefully defined as any failure to meet the contractually stated servicing obligations on time and in full. As Steven Kyle and Jeffrey Sachs (1984) show, many recent partial defaults have taken the form of "rescheduling" of servicing obligations. The present discussion is not concerned either with legal definitions of default or with the use of window dressing, largely induced by regulatory constraints, to avoid classifying debts as legally in default.

In some existing reputational theories of sovereign debt—for example, Jonathan Eaton and Mark Gersovitz (1981) and Kenneth Kletzer (1984)—the desire for continued access to loans can induce a sovereign to service its debts, but these theories do not recognize that sovereign debts actually act like contingent claims and, hence, do not distinguish excusable default from unjustifiable repudiation. Jeremy Bulow and Kenneth Rogoff (1986) develop an alternative analysis of partial default that focuses on the deterrent to repudiation provided by potential infringements of sovereignty, such as the possible seizure by lenders of certain assets of a debtor sovereign that are outside of its jurisdiction and that thereby effectively serve as collateral. In contrast to a model of reputation, this theory deemphasizes the difference between sovereign debt and private debt. In other related literature, Robert E. Lucas, Jr., and Nancy Stokey (1983) develop a formal model of debt as a contingent claim, but they abstract from the problem of enforcing servicing commitments.

I. Analytical Framework

The present analysis considers a sovereign who borrows either to invest in a concave riskfree productive technology that is available only to itself or to purchase riskfree assets that are generally available, and who services debt in such a way as to shift to its lenders the risk associated with an exogenous stochastic component of income. The use of debts that are implicitly contingent claims to facilitate risk shifting presumably results from the desirability of the prepayment of indemnities. By borrowing an amount equal to the maximum indemnity for which it would contract, a large agent like a sovereign who wants to insure itself against the effects of bad states of the world can draw on the resources of many small and anonymous insurers, with whom it would be costly to write and to enforce contracts requiring the payment of an indemnity after the realization of a bad state of the world. The simplifying assumption that the technology is riskfree separates the sovereign's exposure to risk from its decision to invest and enables the analysis to define efficient investment independently of the availability of risk shifting.

Realizations of the exogenous stochastic component of income, denoted by z_t, range from a good state, Z, to a worst state, ζ, and have a stationary probability distribution, $p(z_t)$, with mean $\bar{z}$. The analysis assumes that z_t is verifiable, either directly or indirectly. This assumption is essential to the lenders' ability to treat sovereign debts as contingent claims and to distinguish excusable defaults from unjustifiable repudiation. Examples of important risks that a sovereign faces—such as weather, the terms of trade, external threats, internal discontent—are consistent with this assumption.[2]

The sovereign's objective in period τ is to maximize the expectation of the present value of current and future utility from consumption over a horizon of h periods, which corresponds to the prospective longevity of the sovereign's survival in power. This expectation, U_τ, is given by

$$(1) \quad U_\tau = u(c_\tau) + E_\tau \sum_{t=\tau+1}^{\tau+h} \beta^{t-\tau} u(c_t),$$

with

$$u' > 0,\ u'' < 0,\ u'(0) = \infty, \quad \text{and} \quad 0 < \beta \leq 1,$$

where E_τ is an operator that denotes an expectation conditional on information available in period τ, c_t is an index of consumption in period t, and β is a discount factor that reflects the sovereign's constant rate of pure time preference.

---

[2] In practice, the verification process can involve "negotiations" between the sovereign and its lenders as well as mediation provided by impartial agencies such as the IMF. The prepayment of indemnities as loans with implicitly contingent servicing commitments allows the sovereign to execute the contract without waiting for its lenders to verify the state of the world. If a sovereign also faced important risks whose realizations lenders could not verify, then lenders could confront an additional risk that in some unverifiable states the sovereign would choose not to validate lenders' expectations relating debt servicing to z_t. In Grossman and Van Huyck (1987), we develop a complementary model in which nominally denominated debt facilitates the shifting of risks associated with unverifiable realizations of a contingency that is negatively correlated with the price level.

The analysis assumes that h is a random variable defined over the nonnegative integers, and, for simplicity, assumes that in any period the probability that the current sovereignty will terminate is $1-\gamma$, where $0 \leq \gamma \leq 1$. Thus, the probability that h will turn out to be less than $n+1$ periods is $1-\gamma^n$. Because h is unbounded, evaluating U_τ requires a calculation of expected utility over an infinite horizon, with utility at date t discounted to reflect the probability that h would turn out to be less than t. Specifically, equation (1) implies

$$(2) \quad U_\tau = u(c_\tau) + \sum_{t=\tau+1}^{\infty} (\beta\gamma)^{t-\tau} E_\tau[u(c_t)].$$

The analysis assumes that lenders know β, γ, and the function $u(c)$. Note that, if the probability of termination, $1-\gamma$, is positive, then U_τ is defined even if the discount factor, β, equals unity.

To focus the analysis on the intertemporal calculus that supports the reputational equilibrium, it is useful to abstract from other dynamic considerations that would be associated with the accumulation of either wealth or debt. Accordingly, the analysis assumes that the sovereign does not use either saving or current borrowing to finance current consumption and that the sovereign's debts and investments all mature in one period. These assumptions, together with the assumption that z_t is stationary, mean that the sovereign faces a repeated static problem with a period-by-period budget constraint.[3]

Consumption in period t equals the return from investing the proceeds of last period's borrowing, $F(b_{t-1})$, plus the current stochastic component of income, z_t, minus current debt service, s_t—that is,

$$(3) \quad c_t = F(b_{t-1}) + z_t - s_t,$$

subject to $b_{t-1} \geq 0$ and $s_t \geq 0$.

The form of $F(\cdot)$ reflects the sovereign's choice to invest in its productive technology until the marginal product equals the depreciation rate of unity plus the riskfree interest rate ρ. Accordingly, $F(\cdot)$ satisfies $F' > 1+\rho$ and $F'' < 0$ for $b_{t-1} < B$ and $F' = 1+\rho$ and $F'' = 0$ for $b_{t-1} \geq B$, where B is a nonnegative parameter. (In the special case of B equal to zero, risk shifting is the only motivation for borrowing.) The nonnegativity constraint on s_t reflects the assumption that it is not feasible to write and to enforce contracts that require insurers to indemnify the sovereign after the realization of a bad state of the world.

One possible interpretation on this setup is that c_t represents national consumption and that $F(b_{t-1}) + z_t$ is real national income. Another possibility, which may be more realistic, is to regard the sovereign as a proprietor and to interpret c_t as the proprietor's own consumption and $F(b_{t-1}) + z_t$ as the proprietor's income from taxes and other sources net of expenditures necessary for its survival. Depending on the choice between these interpretations, z_t can mainly reflect either the randomness of factors, like the weather or the terms of trade, that affect real national income or the randomness of factors, like external threats or internal discontent, that affect the net revenues of a proprietary sovereign. In any event, the concavity of $u(c_t)$ motivates the sovereign to issue debt with planned servicing that is contingent on z_t.

The supply price of loans is a critical constraint on the sovereign's choices. The analysis assumes that the sovereign's lenders are atomistic and that market clearing implies that the expected value of debt servicing in period t conditional on information available in period $t-1$ equals the alternative riskfree return on the amount lent to the sovereign in period $t-1$. The expected value of debt servicing is the sum over all possible realizations of z_t of the products of the probability of each state occurring and the amounts of servicing that lenders expect in each state. Accordingly, b_{t-1} cannot exceed an amount given by

---

[3] Although, as explained in fn. 5 below, this setup simplifies the analysis considerably, it does not sacrifice any apparent qualitative generality in the analysis of reputation. The important point is that, given that feasible limits on savings and accumulated debt are finite, these devices, even if available to the sovereign, could not facilitate complete smoothing of consumption. Thus, even if it could vary its wealth over time, the sovereign would want to issue contingent claims to shift residual risk and would have an incentive to maintain its reputation for servicing these claims. Moreover, if it could achieve complete risk shifting, it would not use variations in net wealth to smooth consumption. More generally, in an extended analysis, the randomness of z_t would represent the net variability of resources available for current consumption and debt service after allowing for net current saving.

$$(4) \qquad \sum_z p(z_t) S^e_{t-1}(z_t) = (1+\rho) b_{t-1},$$

where $S^e_{t-1}(z_t)$ is the amount of debt servicing that the lenders in period $t-1$ expect to receive in period t as a function of the realization z_t.

The final component needed to complete the analytical framework is the determination of $S^e_{t-1}(z_t)$. The key element in formulating this expectation is the lenders' perception of the strategy that the sovereign employs in attempting to maximize its expected utility. The different solutions for the properties of sovereign debt derived in the following sections reflect possible differences in these strategies.

II. Efficient Risk Shifting and Excusable Default

To focus the analysis and provide a useful benchmark case, suppose that the sovereign could irrevocably commit itself in period $t-1$ to a state-contingent debt-servicing policy for period t, given by

$$(5) \qquad s_t = S_{t-1}(z_t).$$

This commitment would allow the possibility of excusable default, but would preclude repudiation. By making this commitment, the sovereign would determine the lenders' expectation of actual debt servicing to be

$$(6) \qquad S^e_{t-1}(z_t) = S_{t-1}(z_t).$$

At the same time, the sovereign would have no effective choice regarding the amount of current income to devote to servicing debt issued last period. Its current choices would involve only the amount of new debt to issue and the contingent servicing commitment to attach to this debt.

Because the analysis abstracts from interperiod variations in the sovereign's wealth and assumes that z_t is stationary, the sovereign's choices would be time invariant.[4] Specifically, the maximization of U_τ with an irrevocable state-contingent debt-servicing commitment would reduce to the choice in each period $t-1$ of the value of b_{t-1} and the servicing policy $S_{t-1}(z_t)$ that maximizes $E_{t-1}u(c_t)$, subject to the constraints given by equations (3)–(6). The critical values for this problem, denoted $\tilde{b}$ and $\tilde{S}(z_t)$, would be

$$(7) \qquad \tilde{b} = \max\left(B, \frac{\bar{z}-\zeta}{1+\rho} \right) \quad \text{and}$$

$$(8) \quad \tilde{S}(z_t) = z_t - \bar{z} + (1+\rho)\tilde{b} \geq z_t - \zeta.$$

Equations (7) and (8) indicate that a sovereign who is irrevocably committed not to repudiate would be able both to invest efficiently, thereby maximizing expected consumption, and to achieve efficient risk shifting, which in this case involves shifting to its lenders all risk associated with the stochastic component of income. Equation (7) gives an amount of borrowing $\tilde{b}$ that would be sufficient both for the sovereign to invest in its productive technology until the marginal product equals one plus the riskfree interest rate and for its lenders to prepay the indemnity associated with the worst possible state of the world, which is the discounted value of the difference between $\bar{z}$ and ζ. If B is less than $(\bar{z}-\zeta)/(1+\rho)$, efficient risk shifting requires borrowing beyond the level necessary for efficient investment. (In fact, as Robert Solomon (1981) points out, between

---

[4] If the sovereign could accumulate wealth and debt, and if its discount factor, $\beta\gamma$, differs from the lenders' discount factor, $(1+\rho)^{-1}$, then it would formulate instead a dynamic plan involving a rising or falling expected profile of net borrowing, debt servicing, and expected consumption.

1974 and 1978 non-oil-producing developing countries used almost twenty percent of their net capital inflow to accumulate liquid reserves.)

Equation (8) gives a debt-servicing commitment $\tilde{S}(z_t)$ that calls for adding the difference, which can be positive or negative, between the realization of z_t and the mean value $\bar{z}$ to repayment of loans at interest rate ρ. The interpretation of this servicing policy is that the sovereign would provide full servicing, equal to $Z-\bar{z}+(1+\rho)\tilde{b}$, in the good state, but would provide less than full servicing in inferior states, with the amount of default equal to the difference between Z and z_t. In the case of $\tilde{b}$ equal to $(\bar{z}-\zeta)/(1+\rho)$, the worst state would imply total default. (If $\tilde{S}(z_t)$ were the actual servicing policy, the observation that default, either partial or total, is not the usual outcome would imply that the probability of the good state, $p(Z)$, is large relative to the probability of inferior states.)

With complete risk-shifting consumption, denoted by $\tilde{c}$, would be independent of the realized state of the world and equal to the expected value of income, which is the sum of the net revenue from the sovereign's investments and the mean value of z_t. Specifically, substituting $\tilde{b}$ and $\tilde{S}(z_t)$ into equation (3) gives

$$(9)\quad c_t=\tilde{c}=F(\tilde{b})-(1+\rho)\tilde{b}+\bar{z}\quad \text{for all } t.$$

III. Sovereignty and Repudiation

In reality, sovereigns do not, and by their nature cannot, irrevocably commit themselves to a specific state-contingent debt-servicing policy, or to any other policy. Consequently, although providing a useful benchmark case, the analytical strategy of treating the lenders' expectations of debt servicing as a choice variable for the sovereign, which led to equations (7)–(9), does not provide an empirically relevant analysis of sovereign debt.

To consider another useful benchmark case, suppose that, in addition to being incapable of controlling expectations through irrevocable commitments, the sovereign were to ignore any effect that its current actions have on expectations of its future actions. Instead, suppose that the sovereign were to choose actual current servicing, s_τ, to maximize the expected utility of consumption, U_τ, taking the lenders' expectation of its actual servicing policy for future periods as given. Because the partial effect of reducing s_τ is to increase c_τ and U_τ, the solution to this problem would be to set s_τ equal to zero for all realizations of z_τ. In other words, the solution would be for the sovereign to repudiate any positive servicing commitments.

Assuming that lenders correctly perceive that the sovereign will face the same problem and obtain the same solution in the next period, lenders would anticipate repudiation of all positive servicing commitments in the future. Accordingly, for all $t\geq\tau$, $S^e_{t-1}(z_t)$ would equal zero and by equation (4) b_{t-1} also would equal zero. In other words, the strategy of repudiation, which follows from treating the servicing expectations of lenders as given, would mean that in equilibrium the sovereign would be unable to issue any debt and would be unable to shift any risk. This strategy would be "suboptimal" in the sense that a policy that involved positive debt servicing in some states would yield higher expected utility, if the sovereign could convince its lenders to expect such a policy. The debt-servicing policy $\tilde{S}(z_t)$, given by equation (8), is the best of such policies.

IV. Reputational Equilibrium

The analysis in the preceding sections assumed that the sovereign treats lenders' expectations of debt servicing either as a control variable or as a given. Both theoretical considerations and actual experience suggest that these assumptions fail to capture essential elements of the market for sovereign debt. To develop a more interesting model, suppose that, although the sovereign cannot irrevocably commit itself to a debt-servicing policy and, hence, cannot directly control lenders' expectations about future debt servicing, it can influence these expectations through its choice of the amount of current income to devote to debt servicing. The linkage between current servicing and expectations about future servicing is the sovereign's reputation for debt servicing. Given this linkage, a rational sovereign would consider how its current servicing decision affects its reputation and how its reputation affects its ability to borrow now and in the future. (Assuming that the process by which the sovereign appoints and removes individual policymakers permits the sovereign to translate its objectives into policy, reputation resides with the sovereign and not with individual policymakers.)

An essential property of a reputational equilibrium is that the debt-servicing policy that lenders in period $t-1$ expect the sovereign to follow in period t is self-confirming. This property means that for all possible realizations of z_t the expected present value of the sovereign's utility is at least as large if the sovereign validates the lenders' servicing expectation as it would be if the sovereign were to repudiate its debts. Because the analysis abstracts from interperiod variations in the sovereign's net wealth and assumes that z_t is stationary, the equilibrium debt-servicing policy, denoted by $S^*(z_t)$, is time invariant.

The market-clearing condition, equation (4), associates $S^*(z_t)$ with a time-invariant equilibrium amount of debt, denoted by b^*, according to

$$(10)\qquad \sum_z p(z_t)S^*(z_t)=(1+\rho)b^*.$$

Moreover, the consumption constraint, equation (3), associates $S^*(z_t)$ and b^* with a time-invariant equilibrium relation between consumption and the realization of z_t according to

$$(11)\qquad \begin{aligned} c_t &= C^*(z_t)\\ &= F(b^*)+z_t-S^*(z_t) \quad \text{for all } t. \end{aligned}$$

To analyze the determination of $S^*(z_t)$, and hence of b^* and $C^*(z_t)$, assume that sovereigns always follow a rational debt-servicing policy, except for an infinitesimal fraction, ε, of instances in which sovereigns inexplicably lose the rational ability to resist the temptation to repudiate their debts. A loss of rational restraint could result either from idiosyncratic irrationality or from a breakdown in the process by which the individuals who compose the sovereignty reach their decisions. Either infirmity, however uncommonly it occurs, is intrinsic and irreversible.

Knowing this pattern of sovereign behavior, lenders, when dealing with a specific sovereign, attach probability $1-\varepsilon$, which equals approximately unity, to rational and, hence, trustworthy behavior as long as this sovereign has not repudiated its debts within the memory of these lenders. In this case, lenders would expect the sovereign to follow the debt-servicing policy $S^*(z_t)$, and the sovereign would borrow the amount b^* and obtain consumption $C^*(z_t)$.

If, alternatively, this sovereign has repudiated its debts within the memory of current lenders, these lenders would expect such untrustworthy and irrational behavior in the future. In this case, which occurs with frequency ε, or approximately zero, the market-clearing condition would imply that this sovereign would be unable to issue any debt and, hence, would obtain the autarkic outcome of consumption equal to z_t.

Denote the length of time over which current lenders would remember a repudiation to be k periods, where $0\le k\le\infty$. The analysis assumes that k is a random variable and, for simplicity, assumes that in any period the probability that lenders would permanently forget a past repudiation is a constant $1-\delta$, where $0\le\delta\le1$. Thus, the probability that k would turn out to be less than $n+1$ periods is $1-\delta^n$. If δ equals unity, then k is infinite and a sovereign would never recover a trustworthy reputation once it had been lost.

Given that the sovereign has a trustworthy reputation in period $\tau-1$, these assumptions about lenders' expectations imply that

$$(12)\qquad \begin{aligned} &\text{for } t=\tau,\quad S^e(z_t)=S^*(z_t)\quad\text{and}\\ &\text{for } \tau+h\ge t>\tau,\\ &\qquad \text{either } S^e(z_t)=S^*(z_t)\\ &\qquad \text{if } s_{t-j}=S^e(z_{t-j})\ \text{ for all } j=1,\dots,k\\ &\qquad \text{or } S^e(z_t)=0\quad\text{otherwise.} \end{aligned}$$

To derive the implications of condition (12), define V_τ to be the expected present value of the sovereign's utility over an horizon that corresponds to either the prospective longevity of the sovereign's survival in power or the prospective longevity of lenders' memories of a repudiation, whichever is shorter—that is,

$$(13) \quad V_\tau = u(c_\tau) + E_\tau \sum_{t=\tau+1}^{\tau+\min(h,k)} \beta^{t-\tau} u(c_t).$$

Only at most the next k periods are relevant, because the sovereign's utility after period $\tau + k$ is independent of the decision regarding repudiation in period τ. Given the stochastic processes that generate h and k, equation (13) implies

$$(14) \quad V_\tau = u(c_\tau) + \sum_{t=\tau+1}^{\infty} (\beta\gamma\delta)^{t-\tau} E_\tau[u(c_t)].$$

According to equation (14), the contribution of expected utility in period t to V_τ is smaller the larger is t, the larger is the sovereign's rate of pure time preference, the larger is probability that in any period the current sovereignty will terminate, and the larger is the probability that in any period lenders would forget a past repudiation.

The self-confirming property of the debt-servicing policy $S^*(z_t)$ implies that, together with b^* and $C^*(z_t)$, it satisfies the condition

$$(15) \quad V_\tau^* \geq V_\tau^0 \quad \text{for all possible realizations of } z_\tau,$$

where V_τ^* is the expected utility associated with a program that involves borrowing b^* for all $t \geq \tau - 1$ and servicing debt according to the policy $S^*(z_t)$ for all $t \geq \tau$, and where V_τ^0 is the expected utility associated with borrowing b^* in period $\tau - 1$ and repudiating this debt in period τ. Condition (15) says that for all possible realizations of z_τ a plan that for all $t \geq \tau$ involves validating lenders' expectations that actual servicing will conform to $S^*(z_t)$ would generate in period τ as least as high an expected present value of utility for a rational sovereign as would a decision to repudiate.[5]

To calculate V_τ^*, set c_t in equation (14) for all $t \geq \tau$ equal to $C^*(z_t)$ as given by equation (11). Given the probability distribution $p(z_t)$, this calculation yields

$$(16) \quad V_\tau^* = u[F(b^*) + z_\tau - S^*(z_\tau)] + \frac{\alpha}{1-\alpha} \sum_z p(z_t) \times u[F(b^*) + z_t - S^*(z_t)],$$

where $\alpha \equiv \beta\gamma\delta$ and where equation (10) relates b^* and $S^*(z_t)$. To calculate V_τ^0, observe that by repudiating its debts in period τ, the sovereign would obtain consumption equal to $F(b^*) + z_\tau$.[6] At the same time, according to condition (12) repudiation would cause the sovereign to lose its trustworthy reputation. Consequently, in all subsequent periods, the autarkic equilibrium under anticipated repudiation, analyzed in Section III, would obtain. In this equilibrium, the sovereign is unable to borrow and, accordingly, obtains consumption equal to z_t. Substituting these terms into equation (14) yields

$$(17) \quad V_\tau^0 = u[F(b^*) + z_\tau] + \frac{\alpha}{1-\alpha} \sum_z p(z_t) u(z_t).$$

---

[5] If the analysis did not abstract from interperiod variation in the sovereign's net wealth, then the equilibrium amount of debt and the equilibrium debt-servicing policy would not generally be time invariant. In this case, the analysis would have to calculate instead of V_τ^* the expected value of a dynamic program in which the equilibrium amount of debt and the equilibrium debt-servicing policy in each period would depend on the history of realizations of the stochastic component of income.

[6] If the sovereign could accumulate wealth, then it would consume less than $F(b^*) + z_\tau$ in period τ. It would save part of the amount, $S^*(z_\tau)$, realized from repudiation to use to enhance consumption in future periods that have low realizations of z_t. This possibility would imply a somewhat larger value of V_τ^0 than given by equation (17).

Taking account of reputation, the rational sovereign's problem in period τ is to choose the amount of debt b^*, and the associated implicit contingent debt-servicing policy $S^*(z_t)$, to maximize U_τ as given by equation (2) subject to the constraints given by equations (10) and (11) and condition (15). The fact that the combination b^* and $S^*(z_t)$ satisfies condition (15) ensures that the rational sovereign also chooses in each period to validate $S^*(z_t)$ by actually following the debt-servicing policy given by s_t equal to $S^*(z_t)$. This solution to the sovereign's choice problem implies that together with amount of debt b^* the equilibrium debt-servicing policy $S^*(z_t)$ is the expectation that provides the largest value of U_τ subject to the condition that this expectation is self-confirming. The assumption that the equilibrium corresponds to the best self-confirming expectation is consistent with the sovereign's ability to make an active decision to issue the amount of debt b^* and, thereby, to focus lenders' expectations on the associated debt-servicing policy $S^*(z_t)$.

V. Does Reputation Support Efficient Risk Shifting?

The analysis in Section II derived the contingent debt-servicing policy $\tilde{S}(z_t)$, and the associated levels of debt and consumption, $\tilde{b}$ and $\tilde{c}$, that produced the highest value of U_τ, given $S^e(z_t)$ and s_t equal to $\tilde{S}(z_t)$. If condition (15) is not a binding constraint on the choice of b^* and $S^*(z_t)$, then $\tilde{S}(z_t)$ is a self-confirming debt-servicing policy. Accordingly, $S^*(z_t)$ equals $\tilde{S}(z_t)$, b^* equals $\tilde{b}$, and $C^*(z_t)$ equals $\tilde{c}$. As discussed above, the amount of borrowing $\tilde{b}$ enables the sovereign both to invest efficiently, thereby maximizing expected income, and to achieve efficient risk shifting, which in the present framework involves shifting all risk to the lenders.

If, alternatively, condition (15) is a binding constraint, then $\tilde{S}(z_t)$ is not a self-confirming servicing policy, and lenders would not permit the sovereign to borrow the amount $\tilde{b}$. Such a constraint would prevent the sovereign either from shifting all risk to the lenders or from investing efficiently or both.

To analyze the conditions that determine whether condition (15) is a binding constraint, it is necessary to evaluate condition (15) for $S^*(z_t)$ equal to $\tilde{S}(z_t)$ and b^* equal to $\tilde{b}$. Making the appropriate substitutions in equations (16) and (17) gives

$$(18)\quad V_\tau^*|_{b^*=\tilde{b}} = u\left[F(\tilde{b})-(1+\rho)\tilde{b}+\bar{z}\right] + \frac{\alpha}{1-\alpha}u\left[F(\tilde{b})-(1+\rho)\tilde{b}+\bar{z}\right]$$

and

$$(19)\quad V_\tau^0|_{b^*=\tilde{b}} = u\left[F(\tilde{b})+z_\tau\right] + \frac{\alpha}{1-\alpha}\sum_z p(z_t)u(z_t).$$

Comparison of equations (18) and (19) shows that, if b^* were equal to $\tilde{b}$ and $S^*(z_t)$ were equal to $\tilde{S}(z_t)$, current considerations would favor repudiation because $u[F(\tilde{b})-(1+\rho)\tilde{b}+\bar{z}]$ is smaller than $u[F(\tilde{b})+z_\tau]$, whereas future considerations would discourage repudiation because $[\alpha/(1-\alpha)]u[F(\tilde{b})-(1+\rho)\tilde{b}+\bar{z}]$ is larger than $[(\alpha/(1-\alpha)]\Sigma_z p(z_t)u(z_t)$. Looking more closely at future considerations, the factors that would weigh against repudiation are a larger value of α, greater concavity (risk aversion) of the function $u(\cdot)$, a larger variance of the stochastic variable z_t, and a larger net return from investing the loan $F(\tilde{b})-(1+\rho)\tilde{b}$.

Looking more closely at current considerations, the factors that would weigh against repudiation are a smaller realization of $\tilde{S}(z_\tau)$ and again, greater concavity of $u(\cdot)$. Because the largest value of $\tilde{S}(z_\tau)$ would obtain if the realization of z_τ is the good state Z, the incentive to repudiate would be largest in the good state. Hence, if $\tilde{S}(Z)$ satisfies condition (15), then $\tilde{S}(z_t)$ satisfies condition (15) for all possible realizations of z_t.

Suppose, alternatively, that $\tilde{S}(z_t)$ does not satisfy condition (15) for z_t equal to or greater than $Z-x$, where $0 \leq x \leq Z-\zeta$. In this event, the policy $S^*(z_t)$ would involve less planned servicing for z_t equal to or greater than $Z-x$ than does $\tilde{S}(z_t)$ such that $S^*(z_t)$ satisfies condition (15) as an equality for these states. Accordingly, the equilibrium debt-servicing policy would be kinked with servicing being relatively invariant across realizations of z_t equal to or greater than $Z-x$. Specifically, treating condition (15) as a binding constraint and, hence, as an equality implies that

$$\frac{\partial S^*(z_t)}{\partial z_t} = 1 - \frac{u'[F(b^*)+z_t]}{u'[F(b^*)+z_t - S^*(z_t)]}.$$

This derivative, although positive, is less than unity, in contrast to $\partial\tilde{S}(z_t)/\partial z_t$, which equals unity.

With $S^*(z_t)$ less than $\tilde{S}(z_t)$ for one or more values of z_t, equation (10) implies that b^* is less than $\tilde{b}$. Specifically, equation (10) implies that

$$b^* = \tilde{b} - \sum_{z=Z-x}^{Z} \times p(z)[\tilde{S}(z_t) - S^*(z_t)]/(1+\rho).$$

Given that b^* is less than $\tilde{b}$, equation (7) implies that b^* is less than either B or $(\bar{z}-\zeta)/(1+\rho)$ or both. In other words, the sovereign cannot borrow enough either to invest in its productive technology until the marginal product equals one plus the riskfree interest rate or for its lenders to prepay the indemnity required in the worst state of the world under complete risk shifting or both.

The simplest case to analyze has b^* sufficient for efficient investment but insufficient for efficient risk shifting—that is,

$$\tilde{b} = \frac{\bar{z}-\zeta}{1+\rho} > b^* > B.$$

In this case, because investment remains efficient, the expected value of consumption, given by $F(b^*)-(1+\rho)b^*+\bar{z}$, is still equal to $\tilde{c}$. Actual consumption, as given by equation (11), however, depends on the realized state of the world. Specifically, for realizations of z_t less than $Z-x$, $C^*(z_t)$ is smaller than $\tilde{c}$, whereas, for realizations equal to or larger than $Z-x$, $C^*(z_t)$ is larger than $\tilde{c}$. The amount of borrowing b^* is such that the lenders prepay an indemnity sufficient to equalize consumption for all realizations of z_t less than $Z-x$. Specifically, consumption in these states is given by

$$c_t = \tilde{c} - \sum_{z=Z-x}^{Z} p(z)[\tilde{S}(z_t) - S^*(z_t)] \quad \text{for all } z_t < Z-x.$$

Another instructive case has b^* sufficient for efficient risk shifting but insufficient for efficient investment—that is,

$$\tilde{b} = B > b^* > \frac{\bar{z}-\zeta}{1+\rho}.$$

In this case, the expected value of consumption, given by $F(b^*)-(1+\rho)b^*+\bar{z}$, is less than $\tilde{c}$, the expected value of consumption associated with efficient investment. Moreover, it would not be optimal for the sovereign to choose the servicing policy that would yield the same consumption for all realizations of z_t. A policy of low debt servicing in bad states, which would be required for state-invariant consumption, would reduce borrowing further below the level necessary for efficient investment. If b^* is less than B, the sovereign faces a tradeoff between expected consumption and the variability of consumption. Thus, even though the sovereign is risk averse, $S^*(z_t)$ generally will be such that consumption in bad states is less than in good states.

As a final case, b^* can be insufficient both for efficient investment and for efficient risk shifting—that is,

$$b^* < \min\left(B, \frac{\bar{z}-\zeta}{1+\rho}\right).$$

In this case, the sovereign faces an even less favorable tradeoff between the expected value of consumption and the riskiness of consumption.[7]

---

[7] See Van Huyck (1986) for further analysis of cases in which b^* is less than $\tilde{b}$. As noted above, if, for whatever reason, the level of borrowing is not large enough for complete risk shifting, the sovereign, if it could, would use variations in its net wealth to reduce variations in consumption.

VI. Summary

The power to abrogate commitments without having to answer to higher enforcement authority is the essential property of sovereignty. Consequently, a sovereign's decision to validate lenders' expectations about the servicing of its debts presumably depends mainly on the sovereign's concern for its trustworthy reputation for debt servicing. A trustworthy reputation is valuable because it provides continued access to loans.

This paper has analyzed a reputational equilibrium in a model that interprets sovereign debts as contingent claims that both finance investments and facilitate risk shifting. In this model, lenders differentiate excusable defaults, which are associated with implicitly understood contingencies, from unjustifiable repudiation. The model implies that, although all bad states of the world cause excusable default, only the worst state of the world can produce total default. Moreover, being consistent with lenders' expectations about debt servicing, excusable default does not preclude continued access to loans.

In the reputational equilibrium, the amount of debt and associated debt-servicing expectations are such that the short-run gains from repudiation are smaller than the long-run costs from loss of a trustworthy reputation. Thus, although sovereigns sometimes excusably default, they always resist the temptation, which is greatest in the good state of the world, to repudiate their debts. The equilibrium amount of debt can be sufficient to permit efficient risk shifting and efficient investment or it can be constrained in a way that limits risk shifting and can also restrict investment. The factors that tend to produce such a debt constraint include a high rate of pure time preference for the sovereign, a high probability per period that the sovereign will not survive in power, a high probability per period that lenders would permanently forget a past repudiation, low risk aversion for the sovereign, and a low net return from the sovereign's investments.

REFERENCES

Bulow, Jeremy and Rogoff, Kenneth, "A Constant Recontracting Model of Sovereign Debt," NBER Working Paper No. 2088, December 1986.

Eaton, Jonathan and Gersovitz, Mark, "Debt with Potential Repudiation: Theoretical and Empirical Analysis," *Review of Economic Studies*, April 1981, *48*, 289–309.

Eichengreen, Barry and Portes, Richard, "Debt and Default in the 1930's: Causes and Consequences," *European Economic Review*, June 1986, *30*, 599–640.

Grossman, Herschel and Van Huyck, John, "Nominally Denominated Sovereign Debt, Risk Shifting, and Reputation," NBER Working Paper No. 2259, revised December 1987.

Kletzer, Kenneth M., "Asymmetries of Information and LDC Borrowing with Sovereign Risk," *Economic Journal*, June 1984, *94*, 287–307.

Kyle, Steven and Sachs, Jeffrey, "Developing Country Debt and the Market Value of Large Commercial Banks," NBER Working Paper No. 1470, September 1984.

Lucas, Robert E., Jr., and Stokey, Nancy, "Optimal Fiscal and Monetary Policy in an Economy Without Capital," *Journal of Monetary Economics*, July 1983, *12*, 55–93.

Sachs, Jeffrey, "LDC Debt in the 1980s: Risk and Reform," in *Crises in the Economic and Financial Structure*, P. Wachtel, ed., Lexington, MA: Lexington Books, D.C. Heath, 1982.

Solomon, Robert, "The Debt of Developing Countries: Another Look," *Brookings Papers on Economic Activity*, 2:1981, 593–606.

Van Huyck, John, "Seigniorage, Sovereign Debt, and Reputation," unpublished doctoral dissertation, Brown University, May 1986.

G. E. Moore and J. M. Keynes: A Missing Chapter in the History of the Expected Utility Model

By BRADLEY W. BATEMAN*

In his posthumous essay "My Early Beliefs" John Maynard Keynes outlined his "mental history in the dozen years before the war" and asserted their importance to his later life. Yet, despite Keynes's assertion of the importance of this part of his "mental history" to who he was and what he did in 1938 (the year the essay was written and two years after he published *The General Theory*), there is virtually no treatment in the history of economic thought of this period of his life or of its impact on his economic theory. The purpose of this essay is to treat an episode of that period of his life which has particular *analytical* importance and to suggest some ways in which it might contribute to a clearer understanding of his later work in economics.

The episode in Keynes's early life which the essay examines is his role in G. E. Moore's changing opinion about the importance of rules in ethics. In his classic *Principia Ethica* (1903), Moore had originally asserted that obeying certain rules is a necessary part of ethically correct behavior; but in his *Ethics* (1912) he makes exactly the opposite argument saying that individuals are always justified in judging a case for themselves and are under no obligation to follow general rules of conduct. In the parlance of modern philosophy Moore switched from a rule utilitarian argument to an act utilitarian argument.[1]

The first four sections of the essay detail the role which Keynes played in Moore's change of heart. The final section discusses how the episode affects our understanding of Keynes's work in economics.

I. *Principia Ethica*, Expected Good, and Rules

Principia Ethica was written in an effort to solve two fundamental problems in ethics:

> What kind of things ought to exist for their own sakes? and What kind of actions ought we to perform?
> [1903, p. viii]

G. E. Moore found, however, that at the heart of both questions was the question of the meaning of "good." He reasoned that it is good things which should exist and that the actions we should take are those that bring about the most good. Thus, in order to answer either of the two fundamental questions one must first know what makes something good.

Moore's answer to the question of what is good was shockingly straightforward: good is good. Moore believed that good was "indefinable" in terms of any other concept or thing in the world. In this regard, Moore especially wanted to refute the idea of the utilitarians that good was synonymous with pleasure or utility.

---

*Department of Economics, Grinnell College, Grinnell, IA 50112. More so than most essays, this one depended on the help of many people in many places. I especially wish to thank Michael Halls, Judith Allen, Nancy Finlayson MacIntyre, and D. E. Moggridge. None of them is responsible for errors in the essay. Earlier versions of the essay were presented to the History of Economics Society and the Kress Society. Permission to quote from unpublished manuscripts in the J. M. Keynes' Papers in King's College Library was kindly granted by King's College, Cambridge University. Unpublished writings of J. M. Keynes © The Provost and Scholars of King's College, Cambridge 1970.

[1] J. O. Urmson (1970) is apparently the first discussion of the change in Moore's position. Roy Harrod (1951) discusses the relationship between Moore and Keynes but suggests that it had little or no bearing on Keynes's work in probability. Robert Skidelsky (1983) suggests that there was an intellectual, analytical connection, but does not develop the idea.

> If I am asked 'What is good?' my answer is that good is good, and that is the end of the matter. Or if I am asked 'How is good to be defined?' my answer is that it cannot be defined, and that is all I have to say about it. But disappointing as these answers may appear, they are of the very last importance...if I am right, then nobody can foist upon us such an axiom as that 'Pleasure is the only good' or that 'The good is the desired' on the pretence that this is 'the very meaning of the word.' [1903, pp. 6–7]

Moore accepted that either pleasure or utility *could* be characterized as good, but he did not believe that they *necessarily* were.[2]

Amartya Sen and Bernard Williams (1982, pp. 3–4) have characterized utilitarianism as "welfarist consequentialism," by which they mean that it advocates that actions should be evaluated in terms of their consequences and individual preferences. In contrast to this standard, Moore's system of ethics could be characterized as "objective consequentialism." Moore believed that actions should be evaluated in terms of their consequences, but he believed that the value placed on those consequences was an objectively identifiable amount of good.

Given the existence of the objective quality good, the answers to Moore's two fundamental questions really consisted of making correct ethical judgments of what things possessed this quality and of what actions caused good things to exist. Moore felt, however, that the two types of judgments were frequently confused in philosophical discourse. Making judgments of the first type was a fairly straightforward matter of correct identification and something which posed no serious problems if people could be disabused of the ideas put forward by utilitarians. In the second case, though, there was also the difficult question of identifying causality.

> Whenever we judge that a thing is 'good as a means,' we are making a judgment with regard to its causal relations: we judge *both* that it will have a particular kind of effect, *and* that that effect will be good in itself. But to find causal judgments that are universally true is notoriously a matter of extreme difficulty.... With regard, then, to what are the most frequent objects of ethical judgments, namely actions, it is obvious that we cannot be satisfied that any of our universal causal judgments are true, even in the sense in which scientific laws are so. [1903, p. 22]

Moore's solution to the dilemma was the advice that we follow rules. Certain well-established rules, he felt, represented the accumulated knowledge that in the majority of cases the rule resulted in the highest valued consequence.

Moore's argument in favor of rules is a precursor of the more modern versions of rule utilitarianism in that it depends on a probabilistic argument or an application of what we now term the expected utility (EU) model.[3] Moore's argument was that the best course of action was the one with the highest expected (or most probable) good.

> We have, however, not only to consider the relative goodness of different effects, but also the relative probability of their being attained. A less good, that is more likely to be attained, is to be preferred to a greater, that is less probable, if the difference in probability is great enough to outweigh the difference in goodness. [1903, p. 166]

One must consider *both* the goodness of an action's consequences, and the frequency with which they occur, before deciding how to act; the correct action is then the one which has the greatest frequency of good outcomes and not necessarily the one which has the highest valued (but infrequent) outcomes.

---

[2] Ibid., p. 10. See Bernard Williams (1972, pp. 52–54) for a discussion of the fallacies in Moore's position.

[3] In the case of Moore and Keynes, it is, of course, more correct to refer to the EG (expected good) model. The generic use of EU, regardless of the nature of the utility involved, is established in Paul Schoemaker, (1982, p. 538). For a discussion of the rule and act versions of utilitarianism, including references, see J. J. C. Smart and Bernard Williams (1973, pp. 9–12).

Moore actually felt that *full* knowledge of the probabilities needed for making EU calculations was impossible. One could not know all the possible consequences of one's actions through the infinity of time, or the relative frequencies with which different consequences would occur over such an horizon. For this reason, Moore argued that one must make an *assumption* that the long-run probability of "bad" results from any particular action is very small due to the intervening circumstances beyond one's control; on the basis of this assumption, one could then make short-run EU calculations over the range of time for which one did know the relative frequencies of different consequences.

It was over this shorter, knowable horizon that Moore made his argument for adhering to rules. Moore accepted that most of the rules actually adhered to in society did, in fact, *generally* cause the greatest good.

> It seems, then, that with regard to any rule which is *generally* useful, we may assert that it ought *always* to be observed, not on the ground that in *every* particular case it will be useful, but on the ground that in *any* particular case the probability of its being so is greater than that of our being likely to decide rightly that we have before us an instance of its disutility. In short, though we may be sure that there are cases where the rule should be broken, we can never know which those cases are, and ought, therefore, never to break it.
> [1903, pp. 162–63; italics in original]

II. Keynes's Early Critique of Rules

No one familiar with Keynes's life will fail to understand that as a young man he had little or no regard for general rules of conduct. In "My Early Beliefs," Keynes lists this as one of the chief characteristics of himself and his friends at that time.

> We entirely repudiated a personal liability on us to obey general rules. We claimed the right to judge every individual case on its merits, and the wisdom, experience and self-control to do so successfully. This was a very important part of our faith, violently and aggressively held, and for the outer world it was our most obvious and dangerous characteristic.
> [Keynes, X, p. 446]

Unfortunately for Keynes and his friends they were not only ardently opposed to obeying rules; they were also ardent disciples of G. E. Moore. Having all entered Cambridge at about the time that Moore published *Principia Ethica*, they were greatly influenced by him and became members with Moore in the secret society known as the Apostles. Moore's renunciation of classical (welfarist) utilitarianism had an emancipating effect on the younger Apostles and left them free to contemplate the nature of things which could correctly be characterized as good.

> I went up to Cambridge at Michaelmas 1902, and Moore's *Principia Ethica* came out at the end of my first year. ...its effect on *us*, and the talk which preceded and followed it, dominated, and perhaps still dominate, everything else.... The influence was not only overwhelming;...it was exciting, exhilarating, the beginning of a renaissance.... [Keynes, X, p. 435]

Something in this chemistry was not right, however. Although Moore liberated the younger Apostles from classical utilitarianism he was also an advocate of following rules.

> Now what we got from Moore was by no means entirely what he offered us. ...we set on one side, not only that part of Moore's fifth chapter on 'Ethics in Relation to Conduct' which dealt with the obligation so to act as to produce by causal connection the most probable maximum of eventual good through the whole procession of future ages (a discussion which was indeed

> riddled with fallacies), but also the part which discussed the duty of the individual to obey general rules.
> [Keynes, X, pp. 436, 446]

Unlike the other Apostles, however, who apparently chose just to ignore Moore's argument, Keynes felt compelled to examine its numerous "fallacies" and disprove the argument.

> It was an important object of Moore's book to distinguish between goodness as an attribute of states of mind and rightness as an attribute of actions. He also has a section on the justification of general rules of conduct. The large part played by considerations of probability in his theory of right conduct was, indeed, an important contributory cause to my spending all the leisure of many years on the study of that subject. [Keynes, X, p. 445]

The earliest textual evidence of Keynes's interest in this problem is in a set of notes he wrote in the summer and fall of 1905 entitled "Miscellanea Ethica." There he explores the nuances of Moore's conception of good and outlines a "scheme of a complete ethical treatise." It is here that Keynes first notes an interest in the question of the validity of rules. Under a list of questions to be addressed Keynes includes "the exceptions to the rule that truth stands first" and elsewhere he mentions "the difficult question of the probable grounds of action, and the curious connexion between 'probable' and 'ought.'"[4]

Keynes's first treatment of this "curious connexion" as a problem arising specifically from *Principia Ethica* is in a paper he presented to the Apostles which took its name from the fifth chapter of Moore's book.[5] "Ethics in Relation to Conduct" is an attempt to uncover the fallacies in Moore's argument for rules, and "the crux of the matter lies in the meaning of *probability*."[6]

Keynes's purpose in his paper was to find some way to argue against Moore's probabilistic argument for rules. The tack he chose was to argue against Moore's use of a frequency theory of probability.

> I am not clear as to what interpretation underlies Moore's use of the word [probability]. There is, however, one view at least which does necessitate just such a proof as Moore declares to be necessary; and it is this view which I shall mainly discuss.
>
> Any adequate definition of probability I have never seen, and I am unable to give one; but it is possible to try to refute given definitions and to find out the real questions at issue. It is constantly assumed that the correctness of such an assertion as 'this is probably so' can be confirmed or refuted by future events. In other words it is supposed that the statement of a probability makes a prophecy of certain truths concerning events. According to this view 'x will probably happen' means 'I do not know whether x will happen in any particular case, but if a large number of cases be taken, I do know for certain that x will happen more often than not.' [1905]

Probability had not been defined in *Principia Ethica*, but Keynes's characterization of Moore's implicit theory as a frequency theory was certainly correct. Moore had clearly stated that the argument for rules was based on probability in the sense that certain rules should be followed since they result in the greatest frequency of good outcomes. Similarly, his argument against breaking these rules was the impossibility of knowing in which "particular case" the rule will fail to yield the best possible outcome.

---

[4] All of the quotes in this paragraph are from the unpublished manuscript which is deposited in the Kings College Library, Cambridge University. No page numbers are given since the original manuscript has no pagination.

[5] Robert Skidelsky (1983, p. 152) dates this manuscript 23 January 1904, but there is no date on the manuscript. Keynes makes reference in the paper to Pearson's work, but Keynes does not list having read this work in his reading journal until 1907. There is no mention of Keynes's reading of the paper in the Apostles' records.

[6] All of the references are to the unpublished manuscript which is deposited in the King's College Library, Cambridge University. No page numbers are given since the original manuscript has no pagination.

The type of theory which Keynes found most appealing as an alternative was one in which a probability is a degree of belief in a proposition rather than the relative frequency with which something happens. He argued, for instance, that when one talks about flipping a coin that "the probability of a head is one-half" is *not* a statement which concerns the frequency with which heads have occurred in the past or will occur in the future. Instead, it is a statement of the degree to which we believe that heads will come up given what we know about the coin, the coin tosser, and past experiments of a similar type. (See below, Section III.)

> Surely it is plain that "heads and tails are equally probable" means simply and solely that "there is no known reason favouring either heads or tails" and that this is the meaning of the addition "if the tossing is fair." A toss is clearly fair if no party to the transaction knows of any reason favouring the fall of either heads or tails. And the interpretation that is true in this case seems to me to be equally satisfactory in the general case. A statement of probability always has reference to the available evidence and cannot be refuted or confirmed by subsequent events.

The change to using degrees of belief meant that when one made an EU calculation one was *not* determining the course of action with the highest percentage of good outcomes. Instead, one would be trying to find the action in which one would have the highest degree of belief in a good outcome. With Keynes's theory of probability the good of each possible outcome would be multiplied by the degree of belief that one had in its occurrence, giving a product of good and degree of belief. The summation of these products would be the expected good of an action and might be said to represent the degree of expected goodness of the action.

This alternative conception freed one from Moore's "rules of Common Sense" since one was no longer obligated to act according to a frequency of outcomes. Each case could now be considered on its own merits.

> He [Moore] holds that, in the case of a rule of action which is right generally but not always, an individual is *never* justified in supposing that his is one of the exceptional cases.... All we can say is that a rule of action is in general *probably* right, and this is, as I have tried to show, an entirely different statement [italics in original].

III. *A Treatise on Probability* and Rejecting Rules

Keynes's conception of probability in "Ethics in Relation to Conduct" is, of course, a nascent version of the theory that he propounded in *A Treatise on Probability*.[7] In his book, he denies at much greater length that probabilities are relative frequencies referring not only to Pearson's work but also to that of Ellis and Venn. In the place of the frequency theory he offers his theory that probabilities are "degrees of rational belief."

In his paper, Keynes had suggested that anyone would be capable of knowing the "correct" degree of belief in a proposition, but he had not really explored the idea at length. This "correct" knowledge is of much importance in *Probability*, however. Keynes proposes that the "correct" or "rational" degree of belief refers to a logical relation that exists between a proposition and the evidence supporting the proposition. This logical relationship would have a value of 1 if the proposition's truth were certain on the evidence and a value of 0 if the proposition were false on the evidence. If the evidence did not guarantee the truth or falsehood of the proposition, then there existed a degree of belief between 0 and 1 in the proposition's truth.

> Let our premises consist of any set of propositions h, and our conclusion

---

[7] For an explanation of the theory of probability which Keynes conceived in *Probability*, see Tony Lawson (1985) and Bradley Bateman (1987).

> consist of any set of propositions a, then if a knowledge of h justifies a rational belief in a of degree α, we say that there is a *probability-relation* of degree α between a and h.
> [Keynes, VIII, pp. 4]

In identifying the existence of a logical relationship as the basis for one's degree of belief in the proposition, Keynes very much followed Moore's method of describing the good. Whereas Moore had identified good as an objective concept indefinable in terms of anything else, Keynes also identified probability as an objective concept indefinable in terms of anything else.

> ...in the sense important to logic, probability is not subjective. It is not, that is to say, subject to human caprice. A proposition is not probable because we think it so. When once the facts are given which determine our knowledge, what is probable or improbable in these circumstances has been fixed objectively, and is independent of our opinion. [Keynes, VIII, p. 4]

> A *definition* of probability is not possible, unless it contents us to define degrees of the probability-relation by reference to degrees of rational belief. We cannot analyse the probability-relation in terms of simpler ideas.
> [Keynes, VIII, p. 8]

Employing these two similar conceptions it would then be possible for anyone to make EU calculations solely on the basis of their knowledge of good and probability, without any reference to long-run frequencies of outcomes or "the rules of Common Sense." Keynes devoted Chapter 26 of *Probability* (The Application of Probability to Conduct) to exploring the implications of his fully worked out theory of probability for ethics. He begins by quoting from Moore's argument for rules in *Principia Ethica* and then points out its fallacy.

> This argument seems to be invalid and to depend on a wrong philosophical interpretation of probability.... Mr. Moore's argument must be derived from the empirical or frequency theory of probability, according to which we must know for certain what will happen *generally* (whatever that may mean) before we can assert a probability. [Keynes, VIII, pp. 341–42]

IV. *Ethics* and Moore's Rejection of Rules

Although *Probability* was published in 1921, it was virtually completed by 1914 when Keynes set it aside to begin work for the Treasury. Not until after the war, and the publication of *The Economic Consequences of the Peace* did Keynes return to preparing *Probability* for publication. Even the 1914 date is somewhat misleading, however, for the book is an extension of Keynes's 1908 Fellowship thesis. By then, the conception of probability in Keynes's book was established and his subsequent years of work were devoted largely to clarifying and elaborating his new theory (1908–10) and in developing his theory of induction and the philosophical implications of his theory.[8]

The exact role of G. E. Moore in these clarifications and elaborations is unknown. We know that they consulted about Keynes's work, but the existing correspondence only records their arranging meetings for their discussions on the topic. The impact of Keynes's work on Moore is quite clear, however, in Moore's *Ethics* (1912). Overturning his earlier conclusion that there is a positive role for rules in determining correct behavior, Moore argues in *Ethics* against the idea that rules should always be followed.

Keynes's influence is evident not just in Moore's conclusion, but in the nature of his argument. In both of his statements of the argument against the need to *always* follow rules (pp. 81–82, and 178–181) Moore acknowledges that there are times when following the rule will yield bad results and, thus, that it would be correct not to obey the rule. In examining the *ex ante* grounds for choosing between different courses of action, Moore states the case in the new Keynesian EU terminology.

---

[8] See Skidelsky (1983, pp. 233, 238, 254–56).

> It is admitted that he cannot *know for certain* beforehand which of them will actually have the best results; but let us suppose that he has every reason to think that one of them will produce decidedly better results than any of the others—that all probability is in favour of this view. Can we not say, in such a case, that he absolutely *ought* to choose that one? that he will be acting very *wrongly* if he chooses any other? We certainly *should* actually say so; and many people may be inclined to think that we should be right in saying so, no matter what the results may subsequently prove to be.
> [1912, pp. 193–94, italics in original]

Moore never explicitly defines probability in his second book, but in his numerous uses of the word he is very careful not to use it in the sense criticized by Keynes, that is, to stand for relative frequency.[9] In fact, his first discussion involving probability quite overtly alludes to the conception of probability which Keynes had developed in his Fellowship thesis.

> But even if we cannot be justified either in accepting or rejecting, with absolute certainty, any of the alternative hypotheses which can be suggested, it is, I think, well worthwhile to consider carefully the most important among these rival hypotheses. To realise and distinguish clearly from one another the most important of the different views which may be held about these matters is well worth doing, even if we ought to admit that the best of them has no more than a certain amount of probability in its favour, and that the worst have just a possibility of being true. [1912, pp. 10–11]

Probability very clearly refers in this passage to the degree of belief which one holds in a proposition when one is unsure that the proposition is either true or false. Each of the ensuing references to probability in *Ethics* could only be taken to have this meaning.

The other principal arguments of *Principia Ethica* remained unchanged between 1903 and 1912. He repeats his arguments that there is an indefinable, objective measure of good (which he refers to as intrinsic value) and that actions should be judged in terms of their consequences. His only changes seem to be in the way he uses the word *probability* and in his rejection of rule utilitarianism for act utilitarianism. As Keynes had hoped, the first change had led to the second.

V. Conclusion

There are two types of interpretations of Keynes's work on which this episode bears. One is the recent phenomenon of trying to interpret Keynes's work in economics in light of his earlier work in probability. The other type of interpretation on which the episode bears is a much older one—the idea that *The General Theory* is a form of Walrasian general equilibrium model.

Several commentators have suggested that Keynes's early work in probability was an exercise in developing a subjective theory of probability. Lawrence Boland (1982), for instance, has said "Keynes was famous for his theories of subjective probability" (p. 93) and G. L. S. Shackle (1974), after arguing that Keynes uses a subjective theory of probability in *The General Theory* says that "the pure theory of the matter is from *Probability*." (1974, p. 43) An orthogonal misinterpretation is Allan Meltzer's (1981) argument that Keynes's early work in probability can be characterized as a frequency or stochastic conception of probability.

> Keynes's views on expectations reflect his early writing on probability theory. In this *Treatise on Probability*, and in the *General Theory*, Keynes expresses the view that the best one can do, in an uncertain world, is to act as if the future can be known while recognizing that the important future events are unknown and cannot be foreseen. There is neither conflict nor contradiction when he writes that the long-term

---

[9] See, for example, Moore [1912, pp. 10, 11, 115, 151, 192, 194, 195, 236].

> interest rate and the level of output around which the economy fluctuates are constant, but the future is uncertain and the levels of output and interest rates are subject to violent change. (1981, p. 62)

With a fuller understanding of Keynes's work in probability neither of these types of misinterpretations would be possible. Understanding the symmetry between Moore's conception of good and Keynes's conception of probability should help to reinforce the fact that Keynes's early theory of probability was an *objective* theory. Similarly, an understanding of Keynes's efforts to debunk Moore's argument for following rules should help to clarify that Keynes's early theory was *not* a frequency theory of probability.

Another area of misunderstanding of Keynes's ideas that might be clarified by this essay involves looking at utility. One of the things which an understanding of Moore's influence on Keynes should make clear is that Keynes was not a utilitarian in the same sense as a modern neoclassical economist. One might, of course, argue that this was only true in his early life and that he "came around" later when he took up economics; but it is exactly Keynes's point in his 1938 essay "My Early Beliefs" that no such change had taken place. In his essay, he still credits Moore's greatest influence as his emancipation from Benthamite utilitarianism.

Realizing Keynes's nonutilitarian background helps in understanding his economics in two ways. One way is that it helps to understand that Keynes would not necessarily have been looking for "pareto-optimum" policy results or results that maximized aggregate utility. Since his normative criterion for what constituted "the good" was nonutilitarian there was no reason for him to look for the same desideratum as most modern economists. An example might be his proposals for national self-sufficiency; this policy might give an outcome which could be demonstrated to be "inferior" in a general equilibrium model, but if there is a higher criterion for desirability this result would be inconsequential and immaterial.

The other way that Keynes's nonutilitarian background helps in understanding his economics is in his modeling. If one does not see utility maximizing as the highest goal in life and particularly if one sees oneself as pursuing a higher, objective standard of good, there would be no reason to attribute utility-maximizing behavior to everyone else or to make it the basis for modeling their behavior. Evidence of Keynes's broader horizon in his modeling is found, for instance, in his explanations of saving and consumption behavior in *The General Theory*. After explaining eight reasons why people save money he says, "These eight motives might be called the motives of Precaution, Foresight, Calculation, Improvement, Independence, Enterprise, Pride, and Avarice; and we could also draw up a corresponding list of motives to consumption such as Enjoyment, Shortsightedness, Generosity, Miscalculation, Ostentation, and Extravagance." (Keynes, VII, p. 108) In giving these various motives equal weight in explaining economic behavior, Keynes clearly demonstrates that his modeling is not based exclusively in utility maximizing.

It is perfectly legitimate, of course, to recast Keynes's model in a general equilibrium, utility-maximizing form; but it is an entirely different endeavor to cast Keynes's own thought in this framework. Both Don Patinkin (1976) and Meltzer (1981), for instance, clearly suggest that Keynes's intention was to create a Walrasian general equilibrium model. Patinkin says that, "*The General Theory* can be regarded as the first practical application of the Walrasian theory of general equilibrium," (p. 98) and this idea is implicit throughout Meltzer's paper. On a full understanding of the influence of G. E. Moore on Keynes's early work in probability, however, such interpretations are not tenable. Although Keynes was constructing a model of "general" equilibrium in *The General Theory*, it was not a utilitarian, Walrasian equilibrium.

REFERENCES

Bateman, Bradley, "Keynes's Changing Conception of Probability," *Economics and Philosophy*, April 1987, *3*, 97–119.

Boland, Lawrence, *The Foundations of Economic Method*, London: Allen *R&D* Unwin, 1982.

Harrod, Roy, *The Life of John Maynard Keynes*, New York: Norton, 1951.

Keynes, John Maynard, "Miscellanea Ethica," unpublished manuscript deposited in the King's College Library, Cambridge University, 1905.

______, "Ethics in Relation to Conduct," unpublished manuscript deposited in the King's College Library, Cambridge University, 1905.

______, *The Collected Writings*, D. E. Moggridge and E. Johnson, eds., New York: St. Martin's Press; for the Royal Economic Society, 1971–83.

Lawson, Tony, "Uncertainty and Economic Analysis," *The Economic Journal*, December 1985, *95*, 909–27.

Meltzer, Allan, "Keynes's *General Theory*: A Different Perspective," *Journal of Economic Literature*, March 1981, *19*, 481–517.

Moore, G. E., *Principia Ethica*, Cambridge: Cambridge University Press, 1903.

______, *Ethics*, New York: Henry Holt, 1912.

Patinkin, Don, *Keynes' Monetary Thought*, Durham: Duke University Press, 1976.

Schoemaker, Paul, "The Expected Utility Model: Its Variants, Purposes, Evidence, and Limitations," *Journal of Economic Literature*, June 1982, *20*, pp. 529–63.

Sen, Amartya and Williams, Bernard, *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 1982.

Shackle, G. L. S., *Keynesian Kaleidics*, Edinburgh: Edinburgh University Press, 1974.

Skidelsky, Robert, *Hopes Betrayed*, London: Macmillan, 1983.

Smart, J. J. C. and Williams, Bernard, *Utilitarianism: For and Against*, Cambridge: Cambridge University Press, 1973.

Urmson, J. O., "Moore's Utilitarianism," in Alice Ambrose and Morris Lazerowitz, eds., *G. E. Moore: Essays in Retrospect*, New York: Humanities Press, 1970.

Williams, Bernard, *Morality: An Introduction to Ethics*, New York: Harper & Row, 1972.

Tax Reform and the Stock Market: An Asset Price Approach

By DAVID M. CUTLER*

The incidence of the corporate income tax has been a subject of much dispute. Traditionally, the debate has focused on the importance of general equilibrium effects in evaluating tax burdens. In applied work, incidence analysis has often consisted of estimating the changes in future tax payments from tax reforms. These "cash flow" models of incidence implicitly assume that the relative price effects of tax reforms are unimportant. In general equilibrium analyses, however, these relative price effects, particularly changes in the price of existing capital, are important determinants of the tax burden. "Asset price" models of incidence can consequently yield predictions quite different from those of cash flow models. In theory, the general equilibrium effects associated with the asset price approach are well understood, but little is known about their empirical magnitude.[1]

This paper assesses these asset price effects by examining the stock market's reaction to the Tax Reform Act of 1986. The 1986 Tax Act presents a unique natural experiment for studying asset pricing theories. The Act changed the overall corporate tax burden (a $120 billion increase over five years), the tax rate on corporate income (a decrease from 46 percent to 34 percent), and the relative treatment of old and new capital (the repeal of the investment tax credit and lengthening of depreciation lifetimes). These changes could produce substantial cross-firm heterogeneity in the reaction to tax news.

The paper uses a "differential effects" event study to evaluate the market response.[2] The analysis focuses on two key events in the tax bill's progress: the vote by the House of Representatives for the bill in December 1985, and the vote for a similar bill by the Senate Finance Committee in May 1986. Both events came as surprises to the financial community and are therefore well suited to an event study.

The analysis leads to two conclusions. First, there is some evidence confirming the predictions of the asset price model. After accounting for changes in future cash flows, firms with greater shares of equipment in their capital stocks benefit from the tax change, while firms with greater pre-reform investment rates suffer share declines. These results suggest that the differential taxation of new and old capital could have substantial effects on market values. Second, however, the paper finds little evidence of a large market response to tax news. Excess returns on days containing similar tax news are essentially uncorrelated, and overall market

---

*Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139. I am grateful to Alan Auerbach, Larry Katz, Greg Mankiw, Peter Reiss, two anonymous referees, and especially Martin Feldstein, Jim Poterba, and Lawrence Summers for useful discussions and comments on an earlier draft. Financial support from a National Science Foundation graduate fellowship is gratefully acknowledged.

[1] Formal discussions of the asset price model and its relation to other incidence methods are in Andrew Abel, 1982; Alan Auerbach, 1986; and Lawrence Summers, 1981, 1985; though the distinction between taxes on new and old assets was noted at least by Richard Musgrave: "In the long run, this [tax on new investment] may lead to a decline or increase in the supply of capital, with resulting changes in yield of new assets. These, in turn, give rise to a revaluation of existing assets, and may involve changes in the relative yields and capital values of various assets." [1959, p. 384]. One cash flow model which does allow for many of the general equilibrium effects is Thomas Downs and Patric Hendershott, 1987. Empirical studies of the asset price approach include Andrew Lyons, 1986; and Downs and Hassan Tehranian, 1988.

[2] See Nancy Rose, 1985, for a comparison of this approach with alternative methods of estimating the effects of changes in policy on corporate returns.

reactions to the tax news are small. These results leave unanswered questions about what the tax news meant to the market, and whether the news was efficiently incorporated in stock prices.

The structure of the paper is as follows. Section I discusses the asset price approach to incidence, as well as a competing cash flow approach. Sections II and III detail the methodology employed and identify dates of changing tax expectations. Section IV examines cross-sectional incidence tests. Section V considers more general tests of market movements. Section VI concludes.

I. Tax Reform and Market Value

Two approaches have traditionally been advanced for studying the effects of tax reform on market value. The first is a cash flow approach. Cash flow analyses evaluate tax incidence by estimating changes in after-tax income. While reform-induced demand shifts are sometimes incorporated in these analyses, other general equilibrium effects typically are not. More commonly, the focus is on computing changes in corporate tax payments.

The cash flow approach predicts a substantial market fall in response to the 1986 Act. The House of Representatives' (1986) forecast of corporate tax payments indicates that the Act should raise $84 billion in real terms over its first five years.[3] The effect of this increase on share values should be tempered by the reduced taxation of corporate dividend payments. Estimates by Jerry Hausman and James Poterba (1987) suggest that the marginal tax rate on dividends will fall by 8.1 percent under the new law, which, assuming historical average real dividend growth of 3.4 percent per year, yields a five-year tax savings of $35 billion. The net effect of the reform is thus an increase of $49 billion in the shareholder tax burden, or about 2.5 percent of the end of 1985 value of the stock market.

While the cash flow approach correctly focuses on after-tax returns, it ignores many general equilibrium effects. These are highlighted in the asset price model of incidence. The asset price approach stresses changes in the value of existing assets from tax reform, in addition to the change in future tax payments. Policies that alter future marginal products also affect the value of existing assets through their impact on the supply of capital-intensive goods and thus on product market prices. If these general equilibrium effects are large, they can reverse the partial equilibrium analysis. Consider, for example, the repeal of the investment tax credit. This change reduces the after-tax return to new equipment,[4] but increases the value of existing equipment, which now competes with more expensive (after-tax) new equipment. Higher returns on old capital therefore partly offset the higher taxes that firms must pay on new capital. The net effect on share values is ambiguous, depending on specific technological and adjustment cost assumptions. The asset price approach does, however, offer cross-sectional predictions about the response to tax news. These are detailed in the next section.[5]

II. Methodology and Data

To measure the impact of tax reform on market value, I examine the excess returns caused by tax news. I use the market model to measure these returns:

---

[3] See Alan Auerbach and James Poterba, 1987, for details of the calculation.

[4] Since the investment tax credit applied only to equipment, the tax change hurt equipment investment more than structures investment. Auerbach, 1987, estimates that the tax reform increased the effective tax rate on equipment from −3.3 percent to 38.0 percent, and lowered the rate on structures from 45.6 percent to 37.0 percent.

[5] The 1986 Act also changed the taxation of personal income, which could affect the value of corporate distributions. I consider some of these changes in the empirical work, but concentrate on the corporate tax changes. I also ignore the implications of the asset price model for the transition after the tax news. Since the two events I consider contained similar news, the change in prices after each should be similar; the transitions after the votes, in contrast, need not be. Further, the phasing in of many of the tax changes and the retroactive repeal of the investment tax credit make the determination of the transition path difficult.

$$(1) \qquad R_{it} = \alpha_i + \beta_i R_{mt} + \varepsilon_{it},$$

where R_{it} and R_{mt} are the return to stock i and the market (m) at time t. The coefficient $\beta_i = \text{cov}(R_i, R_m)/\text{var}(R_m)$ is the share of the return that cannot be fully diversified. The excess return is then the residual, ε_{it}.[6] The firm sample is the 1985 *Fortune* 500 largest industrial corporations.[7]

To examine the incidence hypotheses, I model the excess returns as a function of firm asset and financial characteristics:

$$(2) \qquad ERET_i = \tau_0 + \tau_1 M\&ESHARE_i + \tau_2 INV_i + \tau_3 NETK_i + \tau_4 ATR_i + \tau_5 \Delta CASHFLOW_i + \xi_i,$$

where $ERET_i = \varepsilon_{iH} + \varepsilon_{iS}$, the combined excess return from the days or weeks around the House and Senate Finance Committee votes.[8] The first three variables are measures of the firm's assets. *M&ESHARE* is the percentage of the 1985 gross book value of the capital stock accounted for by machinery and equipment (found in the annual report of each company).[9] *INV* and *NETK* are the average growth rate of the net capital stock from 1982 to 1985 and the net book value (in millions of dollars) of the capital stock in 1985. Both variables are from *COMPUSTAT*.

The last two variables highlight the tax payments of the firm. *ATR* is the firm's 1985 average tax rate, while *ΔCASHFLOW* is the projected change in annual tax payments from the tax reform. *ΔCASHFLOW* is found by recomputing the corporation's 1985 average tax rate under 1988 tax rules and multiplying this change by the ratio of the tax base of the firm to the value of equity.[10] The data on average tax rates and changes in average tax rates are from Tax Analysts (1985, 1986).[11] The other variables are from *COMPUSTAT*.

The two incidence hypotheses offer different predictions about the importance of these explanatory variables. The cash flow approach predicts that the projected change in tax payments should be positively related to the firm's excess stock return, but that the other variables should not. After controlling for future taxes, no pre-reform aspects of the corporation should influence share prices.[12]

The asset price approach similarly suggests that the change in cash flow should be positively related to excess returns, since this measures the direct effects of the reform on tax liabilities. The other variables capture the relative price effects. Holding constant future taxes and the stock of existing capital, an increase in the equipment share of the capital stock increases the amount of the firm's existing capital that is in competition with the less subsidized new equipment, and thus should positively affect excess returns. Similarly, the investment rate should be negatively related to excess returns, because faster-growing firms have younger capital stocks and thus generate larger depreciation deductions. The reduced corporate rate decreases the value of these deductions, so that, for equal final capital stocks, higher growth firms should experience larger windfall losses. Finally, the pre-reform average tax rate should have a negative effect on excess returns. Low average tax rates are associated with tax loss firms, or firms nearing tax losses. Since the tax act increases overall corporate tax payments, the value of existing tax losses should rise.[13]

---

[6] I use daily return data from January 1985 to November 1986. The data were adjusted for dividend payments and stock splits. The return on the Standard & Poor's Composite Stock Index was employed as a proxy for the market return.

[7] An industrial corporation is one in which over 50 percent of the sales are from mining or manufacturing. The *Fortune* list was selected because annual reports are more readily available for these firms, and because the companies in the *Fortune* survey are ranked on sales, not profits. The sample thus should not be self-selected for profitability. In fact, a record number of firms (70) lost money in 1985. Due to missing data, the sample was reduced to 336 of the 500 firms.

[8] I report OLS estimates of equation (2), though differential variances of the error terms in equation (1) would result in heteroskedastic errors in equation (2). Using weighted least squares has no significant effects on the estimates.

[9] A more appropriate specification would use the percentage of equipment in the net capital stock. Firms only report asset types by gross value, however.

[10] The tax base is imputed as the sum of federal and foreign tax payments and investment tax credits utilized, divided by the statutory tax rate in 1985 (46 percent).

[11] Tax Analysts did not compute the average tax rate or projected change in the average tax rate for a small number of firms. For these firms, the variables utilized were the weighted average estimates for the relevant industry. The results do not change across samples.

[12] If the tax bases of firms grow at different rates, or if the 1986 and 1987 transition rules differ greatly from the fully implemented provisions, the recalculations of the average tax rates may not be a complete estimate of the total changes in cash flow. While part of these differences may be captured by the industry dummies, the noise in the estimates suggests that the coefficients on the change in cash flow could be downward biased.

III. Expectations of Tax Changes

Two events dominated the history of the 1986 Tax Act. In December 1985, the House of Representatives voted for major tax reform, and in May 1986, the Senate Finance Committee voted for a substantially similar bill. Both of these votes came as surprises to the financial community.[14]

For most of 1985, the tax reform bill was stuck in the House of Representatives. As late as mid-December, the full House had not yet received the bill for consideration. On December 11, two tax bills were introduced on the House floor. The first (Republican) bill was defeated as projected, but in a surprise move, the second (Democratic) bill failed to pass. The headlines on December 12 proclaimed "Reagan Loses Key Tax Vote," and the House adjourned for the weekend "with the prospects growing dimmer that a tax bill can clear the House this year." [*New York Times*, December 13, 1985, IV, 1:3]

Lobbying over tax reform continued that weekend and early the next week. On December 17 (Tuesday), the bill was reintroduced in the House. It was passed later that day, although the vote was not held until 11:00 P.M. The timing of the change in expectations is unclear. It is conceivable that investors anticipated the outcome on Tuesday, though it was not until the discussion of Wednesday's market movement (on Thursday) that the *New York Times* mentioned the bill influencing traders. "Another negative for the stock market was the House passage Tuesday night of its version of the tax-revision bill," it reported. [December 19, 1985, IV, 14:4] Accordingly, I use the return on December 18 (Wednesday) as the immediate response to tax news.

After passing the House, the tax bill was sent to the Senate Finance Committee, where it remained for almost half a year. For most of this period, little progress was made toward approval. As late as May 4, 1986, the *New York Times* noted that the prospects for passage looked dim. Powerful lobbyists were strongly against the bill. After working all day on May 6 (Tuesday), however, the Senate Finance Committee voted on and passed a tax bill. As with the House passage, the vote was not taken until late at night, this time at 12:30 A.M. on May 7. The *New York Times* noted that the outcome was uncertain until a late-night agreement was reached leaving in oil and gas tax shelters. Thursday's paper carried the headline "Tax Revision in 1986 is Almost Assured, Lawmakers Insist," and reported that, "Leading Senators and Representatives from both parties said today (May 7) that the Senate Finance Committee's unanimous approval of a tax bill meant that comprehensive tax revision legislation would almost certainly be enacted by the end of this year." [May 8, 1986, I, 1:6] As with the House vote, I use the day after the vote (May 7) as the immediate response to tax news.

IV. The Effect on Market Values

The most natural test of the reaction to tax news is the change in aggregate share values. The cash flow hypothesis predicts a large decline from the capitalization of the increased taxes, while the predictions of the asset price model are ambiguous. The actual response to the votes was negative but small in magnitude. On the day after the House vote, the S&P 500 fell by 0.40 percent; after the Senate Finance Committee vote, the index fell by 0.49 percent. These responses do not appear to stem from mis-specified dates of expectations changes. On the previous days, the index fell by 0.65 and 0.21 percent. In fact, for the 10-day period spanning the two votes, the index fell by 0.65 percent for the House vote, but actually rose by 1.19 percent for the Senate Finance Committee vote.[15] The aggregate market statistics thus seem inconsistent with large share declines, though the overall movements do little to estimate the relative price effects. The remainder of this section examines excess returns using industry- and firm-specific returns.

---

[13] This effect is partly mitigated by the repeal of the General Utilities doctrine and the strengthening of ownership requirements to utilize acquired operating losses, both of which make the transfer of tax losses more difficult. It is difficult to find proxies to control for these effects, however.

[14] The discussion of the tax bill here is necessarily brief. For a narrative of the bill's history, see Jeffrey Birnbaum and Alan Murray, 1987.

A. *Industry-Based Tests*

Most incidence analyses are conducted on industry aggregates. This is both because demand shifts from the tax reform would plausibly be concentrated along industry lines,[16] and because firms in an industry may be similar enough to merit common treatment. To test these hypotheses, Table 1 presents the average excess returns for the firms in each industry. The industries are grouped into "winners" and "losers" based on newspaper and magazine discussions of the bill's likely effects. The first column reports the predicted changes in one year's net income from one of the quantitative assessments of the tax bill, the Shearson-Lehman Brothers projections (Walter Dolde, 1986, and Allen Sinai, 1986). The Shearson projections are similar to other discussions in the financial press.

The remaining three columns present the industry average return for different time periods: the day after both votes, the 10-day (6-market day) period encompassing the weekend before and after each vote, and the one-month period beginning the Friday before the vote and extending after the vote. The industry returns provide little evidence that reactions to tax reform fell along industry lines. For most industries, one cannot reject the hypothesis of zero excess return. Further, the predicted and actual effects seem only weakly correlated. The industries with above-average predicted changes in income do not appear to have above-average excess returns. In fact, most of the industries in the "losers'" category were consistent winners in the stock market,[17] although the predicted winners fared well on average. The results are generally unsupportive of industry-based measures of tax incidence.

B. *Firm-Specific Tests*

Since industry-level predictions are only partly consistent with the observed reaction, I examine returns on a firm-specific basis. Table 2 presents the results for estimating equation (2). I include industry dummies to control for industry effects and estimate the equations over different time periods to examine the timing of the market reaction.

The estimates in Table 2 lead to three conclusions. The most significant result is the distinction between high growth firms and firms with large existing equipment shares. Exclusive of the one-day return, firms with high investment rates suffer from the tax reform; the effects are significant and increase in magnitude with longer time periods. In contrast, firms with large shares of

---

[15] Other measures of market activity also show little movement. Market volume on the day after both votes (138 million and 130 million shares for the House and Senate Finance Committee votes) was near average for that week (154 and 125 million shares) and only average for that month (137 and 127 million shares). The same is true of the 10-day volume moving average, the number of new highs and lows, and the number of advances and declines.

[16] A common prediction was that the shift in the tax burden from households to corporations would increase demand for consumer goods and decrease demand for corporate equipment.

[17] The oil and gas industry did well after the Senate Finance Committee vote because of the already noted depletion provisions included in the bill. For the other industries, however, special provisions do not appear to be important.

TABLE 1—ESTIMATED AND ACTUAL TAX REFORM WINNERS AND LOSERS[a]

Industry	Predicted Change in Net Income (percent)	Excess Returns Over		
		One Day	Ten Days	One Month
Predicted Winners				
Appliances	8.40	1.35	2.55	6.02
		(1.45)	(3.02)	(5.00)
Apparel	5.10	0.36	2.54	7.13
		(1.02)	(2.14)	(3.54)
Wholesale and Retail Trade	4.50	0.37	0.98	−0.27
		(1.12)	(2.34)	(3.87)
Service	1.80	1.76	1.27	−3.20
		(1.77)	(3.70)	(6.12)
Media	1.10	0.76	1.00	1.70
		(0.72)	(1.51)	(2.50)
Predicted Losers				
Electrical Machinery	−5.65	−0.03	−2.20	−2.03
		(0.54)	(1.12)	(1.85)
Aerospace	−5.82	0.30	−0.30	−2.65
		(0.65)	(1.35)	(2.24)
Machinery and Equipment	−6.39	1.46	3.48	4.29
		(0.50)	(1.05)	(1.77)
Communication	−15.00	−0.38	−0.08	−4.11
		(0.95)	(1.98)	(3.27)
Metals and Mining	−18.50	0.24	−0.96	0.36
		(0.58)	(1.23)	(2.04)
Oil and Gas	−20.30	1.33	3.97	2.98
		(0.46)	(0.26)	(1.61)
Heavy Industry	−76.43	−0.53	0.09	1.09
		(0.76)	(1.58)	(2.61)
Indeterminate				
Dairy Products	14.78	−0.08	−5.60	−6.06
		(1.45)	(3.02)	(5.00)
Stone, Clay, and Plastics	13.12	−0.32	−0.34	−2.38
		(0.67)	(1.40)	(2.31)
Drugs	9.60	−0.86	−1.99	−2.30
		(0.70)	(1.45)	(2.40)
Furniture	8.40	−0.49	−0.23	0.03
		(2.51)	(5.23)	(8.66)
Measuring Instruments	7.83	−1.32	−3.76	2.59
		(0.79)	(1.65)	(2.74)
Food and Tobacco	6.00	−0.18	−0.48	−2.79
		(0.51)	(1.07)	(1.77)
Textiles	5.10	0.44	0.74	7.55
		(0.89)	(1.85)	(3.06)
Beverages	3.40	−0.53	−3.54	−1.37
		(0.95)	(1.98)	(3.54)
Engines and Turbines	1.10	−0.11	2.70	8.89
		(1.12)	(2.34)	(3.87)
Office Machinery	−0.02	0.25	0.27	−0.31
		(0.65)	(1.40)	(2.31)
Automobiles	−13.30	−1.15	−4.39	−3.14
		(0.63)	(1.35)	(2.16)
Miscellaneous Manufacturing	−15.60	−0.97	−0.97	6.16
		(1.12)	(2.34)	(3.87)
Chemicals	−23.20	0.02	0.10	0.10
		(0.45)	(0.94)	(1.56)
Paper and Lumber	−25.80	−0.29	−0.09	1.48
		(0.58)	(1.20)	(2.04)

TABLE 1—CONTINUED

Industry	Predicted Change in Net Income (percent)	Excess Returns Over One Day	Excess Returns Over Ten Days	Excess Returns Over One Month
Rubber	−34.20	−0.16	−3.13	−2.03
		(0.89)	(1.85)	(3.06)
R^2		0.09	0.17	0.12
N		336	333	330

Sources: Stock return data are from Data Resources, Inc. Changes in net income are from Dolde (1986) and Sinai (1986). Periodicals used in grouping industries were the *New York Times*, *Wall Street Journal*, *Fortune*, and *Business Week*.

[a] This table shows predicted change in net income and mean excess return for each industry. Excess returns are the sum of the returns for the House and Senate Finance Committee votes. Industries were grouped by newspaper and magazine discussions of tax reform. Industries in the "Indeterminate" category either had no predictions made about them, or were predicted as a winner by some analysts and a loser by others. Standard errors are in parentheses.

formerly ITC-qualifying equipment increase in value in response to tax news, though the estimates are less significant when industry effects are included. Both of these results support the predictions of the asset price model of incidence.

The equipment share and investment rate coefficients are broadly consistent with the predictions of the asset price theory.[18] The coefficient on the investment rate measures the difference in the losses on existing depreciation allowances for firms with different rates of investment. Using aggregate data, it is possible to provide an estimate of these effects. For a given nominal discount rate (8 percent), it is possible to calculate the present value of depreciation allowances (under the 1985 tax law) for a unit of equipment and structures of every vintage. Given constant depreciation rates, the stock of depreciation deductions on existing assets can then be computed for different rates of investment. Evaluated at the 1982–85 average net investment rates, a 1 percent increase in investment, split between equipment and structures in the ratio of the outstanding stocks (70 percent equipment in the sample) results in a 3.1 percent increase in the stock of depreciation allowances. The rate cut, then, reduces the value of these allowances by 0.37 percent. Empirically, the largest estimate of this effect is about one-third this magnitude. The predicted changes are larger than the observed changes, though allowing for the less favorable tax law before 1981, slower growth in the 1970s, and an effective tax rate change of less than 12 percent would reduce the differences in the stocks of allowances.

The coefficient on the equipment share is more difficult to evaluate because it depends on how quickly the capital stock adjusts to the tax change. A quick adjustment would imply large windfalls from the repeal of the investment tax credit but smaller windfalls from the rate cut. It is possible to provide some estimates of these magnitudes. Auerbach (1986) solves the asset price model analytically and simulates the impact of the 1986 tax reform. He finds that the tax change should result in equal revaluations of equipment and structures when adjustment costs are large, but that each dollar of equipment should increase in value by 0.07 percent more than a dollar of structures when adjustment

---

[18] It is possible that firms with large amounts of net investment suffer because their returns to shareholders are conveyed in the form of capital gains, which are taxed more heavily under the new law. Including the dividend-earnings ratio to measure these payout effects, however, does not change the estimated coefficients. Further, the coefficient on the dividend-earnings ratio is small in magnitude and insignificant statistically.

TABLE 2—TAX CHANGES AND MARKET VALUE[a]

Period	Machinery and Equipment Share	Rate of Investment	Net Capital Stock	Average Tax Rate	Predicted Change In Cash Flow	Industry Dummies	R^2	N
One Day	.018	−.004	.037	−.002	−.045	No	.022	336
	(.012)	(.006)	(.033)	(.004)	(.064)			
One Day	.000	−.003	.032	−.003	−.044	Yes	.097	336
	(.015)	(.006)	(.036)	(.005)	(.066)			
Ten Days	.083	−.029	.068	−.009	−.126	No	.061	333
	(.025)	(.013)	(.070)	(.010)	(.141)			
Ten Days	.035	−.028	.044	−.011	−.166	Yes	.205	333
	(.031)	(.013)	(.075)	(.010)	(.139)			
One Month	.064	−.131	−.025	−.028	−.173	No	.052	330
	(.043)	(.042)	(.117)	(.016)	(.318)			
One Month	.024	−.158	−.027	−.028	.115	Yes	.174	330
	(.051)	(.044)	(.124)	(.016)	(.323)			

Sources: See Table 1 for source of stock returns. Machinery and equipment share is from the footnote of the annual report for each company. Rate of investment and net capital stock are from *COMPUSTAT*. Average tax rates and predicted changes in cash flow are from Tax Analysts (1985, 1986) and *COMPUSTAT*.

[a] This table shows results of OLS estimation of equation (2) in the text. The dependent variable is the sum of the excess returns from the House of Representatives and Senate Finance Committee votes. The independent variables are the share of machinery and equipment in the gross capital stock, the average rate of investment, the net capital stock (in millions of dollars), the pre-reform average tax rate, and the predicted change in cash flow as a percent of total value of equity. The industry dummy variables are the same as in Table 1. Industry estimates are not reported. The final column reports the sample size for the estimate. Firms were omitted from longer period regressions when newspapers reported significant nontax news (generally merger activity) in the period. Standard errors are in parentheses.

costs are small. The estimates of this differential using 10-day and one-month returns (0.03 to 0.08 percent) are thus in line with the Auerbach results.[19]

In contrast to the equipment share and investment rate effects, however, there is no evidence that either average tax rates or predicted changes in future tax payments affect excess returns. The latter result is particularly surprising. Under both approaches, measures of changes in cash flow should be positively related to excess returns; this coefficient is generally negative and insignificant, however. Similarly, the coefficient on the average tax rate provides little evidence for windfall gains from more effective shielding of future liabilities.

These results are quite robust. Replacing the cash flow change with the change in the average tax rate, to eliminate potential tax base problems, does not affect the estimated coefficients. Further, the estimates change little when measures of explicit tax provisions such as minimum tax payments are included in the regression separately or in tandem with the cash flow variable.[20]

---

[19] These estimates suggest two additional points about the asset price model. First the size of the equipment share coefficient implies a much shorter adjustment period than previous studies have found. In the investment function associated with these adjustment costs, Auerbach's 0.07 estimate suggests that the rate of investment increases by 2 percent for a 1-point increase in Tobin's q; the estimates of Summers, 1981, in contrast, are of an increase of 0.031 in the investment rate (see Huntley Schaller, 1987, for larger estimates of the responsiveness to q). Second, the results imply that increases in the value of old capital are consistent with relatively small changes in equilibrium product market prices. With Cobb-Douglas production functions and a capital share of 25 percent, the output of equipment-only firms would have to increase in price by just 1.5 percent more than the output of structures-only firms to produce the 6 percent windfall.

[20] An alternative hypothesis about the tax variables is that, due to incomplete foreign tax credit utilization, firms with large foreign tax payments (essentially min-

TABLE 3—THE CORRELATION OF EXCESS RETURNS[a]

	Dec 16	Dec 17	Dec 18*	Dec 19	Dec 20	Dec 23
May 5	0.115	0.027	0.104	−0.071	−0.002	−0.014
	(0.056)	(0.057)	(0.056)	(0.057)	(0.057)	(0.057)
May 6	−0.045	0.002	0.036	0.176	−0.019	0.036
	(0.057)	(0.057)	(0.057)	(0.055)	(0.057)	(0.057)
May 7*	−0.208	−0.044	0.036	0.124	0.055	−0.071
	(0.054)	(0.057)	(0.057)	(0.056)	(0.057)	(0.056)
May 8	−0.005	−0.022	0.168	−0.003	0.022	0.026
	(0.057)	(0.057)	(0.055)	(0.057)	(0.057)	(0.057)
May 9	0.050	−0.005	0.097	0.087	−0.018	0.021
	(0.057)	(0.057)	(0.056)	(0.056)	(0.057)	(0.057)
May 12	−0.076	0.057	0.121	−0.022	0.036	0.004
	(0.057)	(0.057)	(0.056)	(0.057)	(0.057)	(0.057)

Source: See Table 1.

[a] This table shows correlation of excess returns for the days around the two tax votes. Sample size is 310 firms for all correlations. The excluded firms had missing return data for one or more days. The first day after the vote is indicated with an asterisk. Standard errors are in parentheses.

One final point about Table 2 is troubling as well. The immediate reaction to tax news, shown in the first two rows of the table, cannot be explained by any of the firm characteristics included. All of the estimated coefficients are small in magnitude and statistically insignificant. The same is true when returns on the day before the evening vote are utilized as the dependent variable. At short horizons, returns seem unrelated to the tax news. I take up some implications of this puzzle next.

V. General Tests of the Market Response

The cross-sectional results are consistent with two hypotheses: either tax reform had large effects that are not captured by the included variables; or the news of major tax reform did not significantly affect market valuations. This section explores these possibilities.

As a first test of the importance of tax news, Table 3 presents correlations of excess returns for the two votes. If tax news had a large impact on the market, returns on days of similar news release should be positively correlated. Table 3 provides little evidence of large tax effects. The correlation on the days immediately after the votes is only 0.036, with a standard error of 0.057. This result is striking; on the two most important days in the process of tax reform, stock returns show no indication of common movement. Other days show a larger correlation than this, but there is little evidence of a pervasive reaction to tax news.[21]

Over a 10-day horizon, the results are more promising but still rather weak. The correlation of 10-day excess returns is 0.197, with a standard error of 0.055. While this estimate is statistically significant, it does not indicate when in the period the adjust-

---

ing and oil companies) have much greater tax bases than firms with only U.S. tax liabilities, so that their cash flow projections are understated relative to the other firms. Further, given the large changes in world commodity prices, the effective capital stocks for these firms might be farthest from market value. Re-estimating the equations without these firms, or including the ratio of foreign taxes to total taxes paid as a proxy for these understatements however, does not change the results.

[21] One potential explanation for these results is that the β_i coefficients in equation (1) were changed by the tax reform, and thus the residuals reflect both tax reform news and errors from the change in the covariance. When correlations are estimated using either nominal returns or industry stock returns, where the β_i seem less likely to change greatly, however, the same pattern of correlations results.

ment occurred. Indeed, the within-week correlation of excess returns, like the cross-event correlations in Table 3, are generally very small.

A second test of the response to tax reform is the dispersion of excess returns around the tax votes. If tax reform news was a large part of the market movement, measures of the dispersion of returns on these days should be greater than similar measures on non-news days. This does not appear to be the case, however. The average standard deviation of returns in the weeks of the two votes are 1.81 percent for the House vote and 1.74 percent for the Senate Finance Committee vote. For the 30 days prior to May 1986, a period of no consistently large tax or other news, the average standard deviation is 1.93 percent. Even on the days immediately after the votes, both standard deviations are only 1.96 percent, slightly above the longer period average.

The same finding is true with one-week returns. The Monday-to-Monday return standard deviations are 3.54 percent and 4.01 percent for the two votes, while for the 14 weeks prior to May 1986, the average is 4.41 percent.

There findings are consistent with two possibilities about market expectations of tax changes. Unfortunately, however, neither possibility seems very plausible. First, the events considered might not have changed the expectations of market participants. The long discussions of the tax bill in Congress, or the support for tax reform by President Reagan might have convinced investors that tax reform was soon to occur. Both the immediate attention paid to the votes and the chronicles of journalists who covered the bill's progress (Jeffrey Birnbaum and Alan Murray, 1987), however, make it difficult to believe that tax reform was anticipated long in advance.

Second, the importance of the tax bill might have been reduced because further tax changes were expected after this bill passed. If, for example, investors felt that the bill would be quickly repealed or that tax rates would return to their previous levels soon, the reaction to the tax news would not be large. Evidence for this explanation is found in the statements of congressmen who had hoped to raise revenue from the tax bill, and in the newspaper articles that cautioned against accepting this as the last major tax change. Again, however, the length of the process seems to belie the hypothesis of quick tax revisions.

VI. Conclusions

The stock market response to the 1986 Tax Act suggests a mixed conclusion about the impact of tax reform. First, there is some microeconomic evidence for the predictions of the asset price model of incidence. The revaluation of share prices from the differential taxation of new and old capital and the reduced value of existing depreciation allowances do appear to be empirically important. Second, however, there is little evidence of positive responses to changes in cash flows, or of significant reactions to the tax news more generally. Tests of the variances and covariances of excess returns consistently reject the hypothesis of large, immediate changes in share values. Further, plausible explanations for the small response do not appear to be consistent with the tax reform process.

These latter results especially suggest that an explanation for the observed reaction may be inefficient pricing of the tax news by the market. Recent evidence indicates that changes in macroeconomic fundamentals cannot explain much of the variation in stock returns, and that even large legislative and diplomatic events that are not captured in macroeconomic time-series cannot account for a significant part of the residual variation (David Cutler, Poterba, and Summers, 1988). The limited reaction to the tax news could thus be indicative of this more general inexplicability of stock returns with economic fundamentals.

REFERENCES

Abel, Andrew, "Temporary and Permanent Tax Changes in a *q* Model of Investment," *Journal of Monetary Economics*, March 1982, *9*, 353–73.

Auerbach, Alan J., "Tax Reform and Adjustment Costs: The Impact on Investment and Market Value," NBER Working Paper No. 2103, 1986.

______, (1987a) "The Tax Reform Act of 1986 and the Cost of Capital," *Journal of Economic Perspectives*, Summer 1987, *1*, 73–86.

______ **and Poterba, James M.**, (1987b) "Why Have Corporate Tax Revenues Declined?," in Lawrence H. Summers, ed., *Tax Policy and the Economy*, Cambridge: MIT Press, 1987, 1–28.

Birnbaum, Jeffrey H. and Murray, Alan S., *Showdown at Gucci Gulch: Lawmakers, Lobbyists, and the Unlikely Triumph of Tax Reform*, New York: Random House, 1987.

Cutler, David M., Poterba, James M. and Summers, Lawrence H., "What Moves Stock Prices?," NBER Working Paper No. 2538, March 1988.

Dolde, Walter, "The Impact of Corporate Tax Reform by S&P Industry," *Shearson-Lehman Brothers Prospects*, May 23, 1986.

Downs, Thomas and Hendershott, Patric H., "Tax Policy and Stock Prices," *National Tax Journal*, June 1987, *40*, 183–90.

______ **and Tehranian, Hassan**, "Predicting Stock Price Responses to Tax Policy Changes," *American Economic Review*, December 1988, *78*, 1118–30.

Gordon, Roger H., Hines, James R., Jr. and Summers, Lawrence H., "Notes on the Tax Treatment of Structures," in Martin Feldstein, ed., *The Effects of Taxation on Capital Accumulation*, Chicago: University of Chicago Press, 1986, 223–54.

Hausman, Jerry A. and Poterba, James M., "Household Behavior and the Tax Reform Act of 1986," *Journal of Economic Perspectives*, Summer 1987, *1*, 101–19.

Lyons, Andrew, "How Do Stock Prices Change When Corporate Tax Laws Change? An Empirical Analysis of the Effect of Tax Laws on the Value of the Firm," Princeton University, unpublished doctoral dissertation, 1986.

Musgrave, Richard A., *The Theory of Public Finance*, New York: McGraw-Hill, 1959.

Poterba, James M., "Explaining the Yield Spread Between Taxable and Tax-Exempt Bonds: The Role of Expected Tax Policy," in Harvey S. Rosen, ed., *Studies in State and Local Public Finance*, Chicago: University of Chicago Press, 1986, 5–49.

Rose, Nancy L., "The Incidence of Regulatory Rents in the Motor Carrier Industry," *Rand Journal of Economics*, Autumn 1985, *16*, 299–318.

Schaller, Huntley, "Investment and Acquisitions: The Evidence from U.S. Firm Data," MIT mimeo., November 1987.

Sinai, Allen, "The Tax Reform Act of 1986: Winners and Losers," *Shearson Lehman Brothers Economic Studies Series*, Report No. 22, 1986.

Summers, Lawrence H., "Taxation and Corporate Investment: A *q*-Theory Approach," *Brookings Papers on Economic Activity*, 1:1981, 67–140.

______, "The Asset Price Approach to the Analysis of Capital Income Taxation," in George R. Felwil, ed., *Issues in Contemporary Macroeconomics and Distribution*, London: Macmillan, 1985, 429–43.

Tax Analysts, *Effective Corporate Tax Rates*, Arlington, VA, 1985.

______, *Quantifying the Impact of the Tax Reform Act of 1986 on Effective Corporate Tax Rates*, Arlington, VA, 1986.

U.S. Congress, House of Representatives, *Tax Reform Act of 1986: Conference Report to Accompany H.R. 3838*, Report 99-841, Washington: USGPO, 1986.

Predicting Stock Price Responses to Tax Policy Changes

By THOMAS W. DOWNS AND HASSAN TEHRANIAN*

Events leading to stock price windfalls happen for a variety of reasons. Lately huge gains and losses have resulted from changes in federal tax policies. Alan Auerbach and Laurence Kotlikoff (1983) estimate the Economic Recovery Tax Act (ERTA) of 1981 induced a $200 to $300 billion capital loss. More recently, Lawrence Goulder and Lawrence Summers (1986) and Thomas Downs and Patric Hendershott (1987) estimate that the Tax Reform Act of 1986 provided corporate shareholders with a 12 or 13 percent windfall gain. The legislative granting of windfall gains or losses gives rise to an important question of public policy: What is the impact of major tax changes on corporate asset values?

The objective of this study is to introduce and test a model for estimating windfalls resulting from tax policy changes. The model describes fundamental equity value as a function of capital accumulation patterns, tax policy parameters, and the equilibrium between rental prices and the expected return on capital. The impact on fundamental equity value of the policy changes enacted with ERTA is simulated and those results are used as predictions about stock price windfalls. Such windfalls are computed for different aggregates and actual stock returns data are analyzed to see whether the predicted windfalls accrued.

Section I introduces the valuation model. Section II presents estimates of fundamental equity value for the U.S. nonfinancial corporate sector and three manufacturing industries. Section III simulates the change in fundamental equity value from ERTA, and Section IV analyzes stock returns data to determine whether predicted price adjustments occurred. Section V is the summary.

I. The Valuation Model

Fundamental equity value is the present value of the cash flows expected to accrue to shareholders discounted by the equity financing rate. Denote as $R^j_{s,t}$ the expectation formed at time s about the equity residual cash flow to be received at time t from productive factor j. The fundamental equity value of shareholder claims against productive factor j, denoted V^j_s, is

$$(1) \qquad V^j_s = \sum_{t=s+1}^{\infty} (1+e)^{-(t-s)} R^j_{s,t},$$

where e denotes the equity financing rate. Computing V^j_s requires specification of the equity financing rate and the equity cash flow stream for each factor of production.

The equity financing rate can be determined from a portfolio equilibrium. Letting τ^e denote the effective personal tax rate on equity returns and τ^i the effective tax rate on interest income, then

$$(2) \quad (1-\tau^e)e = (1-\tau^i)\rho + (1-\tau^e)\beta\gamma^e.$$

ρ is the pretax riskfree interest rate, γ^e is the pretax risk premium for interest in the market portfolio of corporate equities, and β is the proportion of premium paid by the producer. The equity financing rate can be determined from equation (2) given exogenous estimates for τ^e, τ^i, ρ, β, and γ^e.

Equity cash flows are generated by only capital factors of production because noncapital inputs do not pass cash flows to equity. Say capital type j is expected to provide a pretax return (earnings before tax depreciation, interest, and taxes) equal to NOI^j and promises tax depreciation deductions equal to $TAXD^j$. Given financing by debt and equity, the cash flow for equity equals after-tax NOIs plus the value of the depreciation tax shield less after-tax interest and debt repayments,

---

*Finance Department, Boston College, Chestnut Hill, MA 02167. We are especially grateful to Patric Hendershott for direction and to George Aragon, Robyn McLaughlin, Nickolaos Travlos, and Jerry Viscione for discussions. Funding has been provided by Boston College Research Grants programs. Anonymous referees have provided generous help. Errors are ours.

$$(3)\quad R^j_{s,t} = (1-\tau)NOI^j_t + \tau TAXD^j_t - (1-\tau)INT^j_t + \Delta DEBT^j_t.$$

τ is the corporate profits tax rate and INT^j and $\Delta DEBT^j$ are the interest and principal cost of debt services associated with financing capital type j. V^j_s is

$$(4)\quad V^j_s = \sum_{t=s+1}^{\infty} (1+e)^{-(t-s)} \times \left[(1-\tau)NOI^j_t + \tau TAXD^j_t - (1-\tau)INT^j_t + \Delta DEBT^j_t\right].$$

Later in this study fundamental equity value is estimated from equation (4). The remainder of Section I describes the modeling of INT, $\Delta DEBT$, NOI, and $TAXD$.

Capital market equilibrium imposes a restriction on the values of all financial claims (Franco Modigliani and Merton Miller, 1963). Total capital income is $(1-\tau)NOI + \tau TAXD$ and in equilibrium it equals the after-corporate-tax cost of financing. The fundamental value of financial claims on capital type j at time t, denoted W^j_t, is

$$(5)\quad W^j_t = \sum_{u=t+1}^{\infty} (1+r)^{-(u+t)} \times \left[(1-\tau)NOI^j_u + \tau TAXD^j_u\right].$$

r is the weighted average after-tax cost of financing, $\alpha(1-\tau)i + (1-\alpha)e$, where i is the pretax debt financing rate and α is the debt-to-market value ratio. The interest and principal cost of debt services are

$$(6)\quad INT^j_t = i\alpha W^j_{t-1}, \quad \text{and}$$

$$(7)\quad \Delta DEBT^j_t = \alpha\left(W^j_t - W^j_{t-1}\right).$$

V^j_s measures the discounted value of the cash flows expected from existing capital plus the net present value of future investments. Previous studies indicate that in aggregate measurements the net present value of future investments is very small. Daniel Holland and Stewart Myers (1981) provide estimates that in the nonfinancial corporate sector in 1978 the net present value of future investments is less than one-half of 1 percent of the capital stock's replacement cost. Estimates by William Brainard, John Shoven, and Laurence Weiss (1980), imply that in 1977 the net present value of future investments for 187 large industrial firms is −1.39 percent of market value and between 1968–77 the average is 1.59 percent. Because the net present value of future investments as a proportion of existing capital is relatively small, NOI_t and $TAXD_t$ are assumed to result solely from capital in-place at time s.

Insights about the NOI stream can be made by analyzing the marginal investment equilibrium of zero net present value. As discussed by Thomas Downs (1986), the replacement cost of a unit of new type j capital, denoted q^j_s, is at equilibrium with its discounted income stream when

$$q^j_s = \nu^j q^j_s + \sum_{t=s+1}^{\infty} (1+r-\pi)^{-(t-s)} \times (1-\tau)c^j_s\left[1 - \sum_{u=1}^{t-s} h^j_u\right] + \tau q^j_s Z^j.$$

ν is the effective rate of the investment tax credit, π is the rate of expected inflation, c_s is the time s rental price of capital, and Z is the present value of depreciation deductions per dollar of investment. The percent of original productive capacity lost by capital u periods after its acquisition is h_u. For example, with geometric depreciation $h_u=\delta(1-\delta)^{u-1}$; with straight-line depreciation over a service life of L years, $h_u=1/L$ for $u=1,\dots,L$ and $h_u=0$ for $u>L$. The term in square brackets $[\cdot]$ is the proportion of time s investments surviving until time t.

Isolating the rental price from the zero net present value equilibrium shows

$$(8)\qquad c_s^j=q_s^j(r-\pi)(1-\nu^j-\tau Z^j)/\left[(1-\tau)(1-H^j)\right].$$

H is a depreciation term, $0\leq H<1$, computed as

$$(9)\qquad H^j=\sum_{u=1}^{\infty}(1+r-\pi)^{-u}h_u^j.$$

For nondepreciating, infinite life capital, H is zero, and for assets that depreciate rapidly, H is close to unity. When assets depreciate over an infinite life at the geometric rate δ, H computed from equation (9) is $\delta/(r+\delta-\pi)$ and the rental price computed from equation (8) is $q_s(r+\delta-\pi)(1-\nu-\tau Z)/(1-\tau)$, the formulation introduced by Robert Hall and Dale Jorgenson (1967).

The rental price is the NOI earned by a unit of capital. When adjustment costs are absent or equal for assets of different vintage the NOI on one unit of new capital is the same as that on old capital. In the presence of adjustment costs that vary across assets of different vintage the NOI for new and old capital differ. Herein, the assumption is maintained that adjustment costs are zero for all capital and they are unaffected by tax policy changes.[1]

Total NOI^j for the capital stock equals the product of rental price times real capital stock. Given exogenous economic depreciation,[2] the stock of real type j capital in-place at time s that survives until time t is

$$(10)\qquad K_t^j=\sum_{u=t-s}^{\infty}I_{t-u}^j\left[1-\sum_{v=1}^{t-s}h_v^j\right].$$

I^j denotes real investment. Because the rental price rises with inflation, total NOI expected at time t is

$$(11)\qquad NOI_t^j=c_s^j(1+\pi)^{(t-s)}K_t^j.$$

Future tax deductions promised by the existing stock, $TAXD^j$, are predetermined by historic investment flows and the tax depreciation laws in effect at the time the capital was put in-place. Let $z_{s,u}^j$ be the proportion of time s gross investments deductible for tax purposes at time $s+u$. The tax depreciation deduction that the time s capital stock promises at t is

$$(12)\qquad TAXD_t^j=\sum_{u=t-s}^{\infty}q_{t-u}^jI_{t-u}^jz_{t-u,u}^j.$$

Equations (11) and (12) specify the construction of return streams spanning times $t=s,\dots,\infty$ that may be used to find the present value of the capital income stream.

II. Fundamental Equity Value Before ERTA

Estimation of fundamental equity value requires construction and substitution of equations (5) through (12) into (4) for each type of capital asset. According to the U.S. Federal Reserve Board (FRB, 1985) Current Cost Balance Sheets, the \$3,773 billion of total assets in the nonfinancial corporate sector (NFC) at year-end 1980 are comprised of fixed assets (47 percent), short-term assets (43 percent), and land (10 percent). For short-term assets and land fundamental values are assumed equal to the FRB entries, but for fixed assets the valuation model introduced above is employed. Separate streams and values for plant and equipment are estimated because of differences in service lives and tax treatment but the discussion below usually refers to the sum or weighted average of the two streams. Discussion focuses first on the NFC exclusively; industry estimates are described toward the end of Section II.

---

[1] Auerbach (1986) shows that in the presence of adjustment costs the windfall is accentuated because of the transient inequality between the return on capital and rental prices.

[2] Hall and Jorgenson (1967) assume exogenous depreciation along a geometric pattern, and Charles Hulten and Frank Wykoff (1981) provide supporting evidence. The geometric assumption has been challenged by Robert Eisner (1972) and Martin Feldstein and Michael Rothschild (1974). Capital stock estimates generated by the U.S. Bureau of Economic Analysis (1986) as do our estimates herein, assume depreciation is exogenous along a straight-line pattern.

Our estimate of the equity financing rate (e) is 0.1708 and is constructed in accordance to equation (2) with the following settings. Debt rates hovered in the 13 percent range during the period under study so ρ is set to 0.13 and the equity risk premium (γ^e) is set to 0.07, the mean annual premium on common stocks between 1948–83 (Roger Ibbotson, 1984). The tax rate on interest income (τ^i) is 0.366 and is computed by Joe Peek and James Wilcox (1983) as the average marginal tax rate on interest income. The tax rate on equity returns (τ^e) is 0.182 and is computed as a weighted average of tax rates on dividends and capital gains.[3] β for the NFC is set to unity.

The length of the cash flow streams equal the average service life of capital, 30 and 13 years for plant and equipment, respectively, based on data from the Bureau of Economic Analysis (BEA, 1986).[4] Column 1 of Table 1 lists the proportion of the year-end 1980 real capital stock remaining productive each year into the future. By the seventh year, 1987, the surviving stock is capable of producing less than half of its 1981 product. After the thirteenth year the stock of equipment is totally expired and after the thirtieth year, 2010, all remnants of the year-end 1980 real capital stock have expired.

The NOI equals the product of the real capital stock times the rental price. A dollar of equipment generates \$0.2347 of NOI whereas per dollar of plant it is \$0.1789.[5] The NOI stream for plant and equipment is listed in column 2. One-third of all NOIs during the first six years of the return stream are free from taxation because of the shield provided by depreciation deductions. Estimates of the tax shield that assets existing at year-end 1980 were expected to provide over their remaining lives are obtained by depreciating historic investments with the tax depreciation practices in effect at their time of acquisition, as specified in equation (12). The proportion of new investments (plant or equipment) depreciated by accelerated methods and tax lives are taken from the SSRC-MIT-PENN Quarterly Econometric Model (1983). Half of all investments depreciated by accelerated methods are depreciated by sum-of-year's digits and half by 200 percent declining balance (150 percent for plant after 1969) with an optimal switch to straight-line. Investments not depreciated by accelerated methods are depreciated by straight-line. The stream of tax depreciation deductions promised by the year-end 1980 capital stock is listed in column 3 of Table 1.

The fundamental value of debt and equity claims on plant and equipment, denoted W(P&E), can be computed by discounting the NOI and $TAXD$ streams with the weighted average cost of financing. Discounted after-tax NOIs plus discounted depreciation tax savings yield a W(P&E) of \$1,218 billion (see equation (5) and col-

---

[3] Joe Peek supplied generous discussion and data about τ^i. For τ^e, the weight applied to the dividend tax rate is 0.40. Following Hendershott (1986) the dividend tax rate is $\tau^{\max}/2$ and the capital gains tax rate is $(1\text{-exclu})\tau^{\max}/4$, where $\tau^{\max}$ is the statutory maximum personal tax rate (70 percent pre-ERTA and 50 percent post-ERTA) and "exclu" is the capital gains exclusion of 0.60. The divisions by 2 and 4 reflect deferral and avoidance activities.

[4] John Musgrave and Kenneth Rogers at the Department of Commerce (Bureau of Economic Analysis and Office of Business Analysis, respectively) have provided generous discussion and data. Our estimate of NFC plant and equipment at year-end 1980 is \$1,640 billion and is comparable to the BEA estimate of \$1,697 billion.

[5] The rental price constructed according to equation (8) assumes: ν is 0.0610 for equipment and 0.0 for plant (N. Behravesh, 1985); τ is 0.4924 reflecting the marginal tax rate of 0.46 and a deductible state and local tax rate of 0.06; π is 0.08; Z is 0.5997 for equipment and 0.3182 for plant and is constructed from the depreciation schedules described in the text; r is 0.1362 based on an e of 0.1708, i of 0.1300, α of 0.33; H is 0.6964 for equipment and 0.4782 for plant.

TABLE 1—FUTURE SERVICES FROM THE 1981 CAPITAL STOCK

Year	Capital Surviving (1)	NOI_t (2)	$TAXD_t$ (3)	$(1-\tau)NOI + \tau TAXD$ (4)	$R_{1980,t}$ (5)
1981	100.0 percent	\$338.5	\$133.6	\$237.7	\$182.4
1982	89.9	328.4	128.6	230.0	175.5
1983	80.4	316.5	110.4	215.0	163.6
1984	71.3	303.0	91.3	198.8	151.1
1985	62.8	287.7	78.3	184.6	140.0
1986	54.9	270.7	68.2	171.0	129.4
1987	47.5	252.6	60.7	158.1	119.4
1988	40.9	234.0	53.3	145.0	109.3
1989	34.9	214.9	46.1	131.8	99.2
1990	29.5	195.5	38.4	118.1	89.0
1991	24.8	176.7	17.1	98.1	74.8
1992	20.8	159.9	16.2	89.2	68.1
1993	17.8	146.7	15.4	82.0	62.6
1994	15.5	138.3	14.6	77.4	58.9
1995	13.8	132.5	13.9	74.1	56.0
1996	12.1	125.8	13.1	70.3	52.8
1997	10.5	118.3	12.4	66.2	49.4
1998	9.1	110.2	11.7	61.7	45.7
1999	7.7	101.4	11.0	56.9	41.9
2000	6.5	92.0	10.2	51.8	37.9
2001	5.4	82.2	9.5	46.4	33.7
2002	4.4	72.0	8.8	40.9	29.5
2003	3.5	61.5	8.1	35.2	25.2
2004	2.6	50.9	7.5	29.5	20.9
2005	2.0	40.6	6.6	23.9	16.8
2006	1.4	30.8	5.8	18.5	12.9
2007	0.9	21.6	5.0	13.4	9.3
2008	0.5	13.3	4.3	8.9	6.1
2009	0.2	6.5	3.7	5.1	3.5
2010	0.1	1.8	3.0	2.4	1.6
Discounted Sum with $r=13.62$ percent		\$1,869	\$548	\$1,218	
Discounted Sum with $e=17.08$ percent					\$816

Notes: All dollars are in billions and refer to the stock of nonfinancial corporate plant and equipment in-place at the beginning of 1981. Estimates are constructed from formulas and data described in text.

umn 4)). Fundamental equity value, denoted V(P&E), is $(1-\alpha)$ times that or \$816 billion. The identical V(P&E) can be obtained by constructing the INT and $\Delta DEBT$ streams and discounting the equity residual cash flow stream by the equity financing rate (see equation (4) and column 5)).

The fundamental value of total assets equals W(P&E) plus FRB short-term assets and land and is \$3,294 billion. It exceeds the \$2,683 billion value of debt and equity claims, meaning that either securities are "undervalued" or that the fundamental value estimate is too high. The ratio of total market value to total fundmental value, referred to herein as the "overvalue ratio," is 0.81 for the NFC.

In the same fashion estimates of V(P&E) are generated in the Food Products (SIC 20), Paper and Allied Products (SIC 26), and the Stone, Clay, and Glass (SIC 30) industries.[6]

[6] The BEA data and our fundamental value estimates are on an establishment definition. The *COMPUSTAT* and *CRSP* data are on an enterprise basis. Our fundamental value estimates are multiplied by the year-end 1980 ratio of "*COMPUSTAT* Industry Sample Total

TABLE 2—FUNDAMENTAL EQUITY VALUE OF FIXED ASSETS AND PREDICTED EQUITY WINDFALLS

Total Nonfinancial Corporate	Food Products SIC 20	Paper Products SIC 26	Stone, Clay, and Glass SIC 30
1 Before-ERTA			
$816.3	$13,533	$11,714	$2,571
2 Constant Personal Tax and Interest Rates but Full Adjustment to ACRS			
$725.4	$11,675	$10,127	$2,039
−6.98 percent	−4.88 percent	−6.50 percent	−9.90 percent
3 Constant Interest Rates but Full Adjustment to ACRS and Personal Taxes			
$724.2	$11,662	$10,114	$2,041
−7.07 percent	−4.92 percent	−6.56 percent	−9.86 percent
4 Full Adjustment to ACRS, Personal Tax, and Interest Rates			
$737.2	$11,908	$10,278	$2,108
−6.07 percent	−4.27 percent	−5.89 percent	−8.63 percent

Notes: Dollars are in billions for the NFC and millions for the industries. The percentages are the predicted excess rate of return to shareholders, computed as the change from Before-ERTA fundamental value relative to equity market values. Equity market value is 1,303 billion dollars in the NFC and in SICs 20, 26, and 30 they are 38,059, 24,399, and 5,372 million dollars, respectively. Estimates are constructed from formulas and data described in text.

The same interest and income tax rates are used in all industries but each industry has its unique equity beta (0.796, 0.760, and 0.992 in SICs 20, 26, and 30, respectively), debt ratio[7] (0.4920, 0.3943, and 0.6840, respectively), tax lives (Robert Coen, 1975), effective rate of the investment tax credit (N. Behravesh, 1985), asset service lives (equipment: 14, 16, and 15 years; plant: 28 years), and capital investment data (BEA). The fundamental equity value of plant and equipment is $13.5, $11.7, and $2.6 billion in SICs 20, 26, and 30, respectively. The overvalue ratios are 0.98, 1.09, and 0.75.[8] The market values of debt and equity in Food and Paper Products are reconcilable to within 10 percent of their fundamental values, but Stone, Clay, and Glass appears undervalued by a quarter.

III. Fundamental Equity Value After ERTA

The Economic Recovery Tax Act of 1981 affected stock prices through several channels. First is the influence of changes in the treatment of tax depreciation deductions on new investments. Second is the influence of changes in personal tax rates and third is the influence of changing interest rates because of shifts in the demand for capital. These influences are considered separately and are empirically summarized in Table 2.

Net Fixed Assets" to "BEA 2-digit SIC Industry Historic Cost Replacement Cost," 0.8580, 0.8248, and 0.7120 for SICs 20, 26, and 30 so that all industry estimates in this study are stated on an enterprise basis.

[7] The calculation of the debt to market value ratios, α, follows the procedure used by George VonFurstenberg, Burton Malkiel, and Harry Watson (1980) and employs market to book value ratios for long-term debt of 0.6665, 0.7332, and 0.6337 in SICs 20, 26, and 30, respectively.

[8] In SIC 20 W(P&E) is $26,640 million. The total fundamental value of all other assets is $49,779 million and is computed from the *COMPUSTAT* (1984) industry sample as "Total Assets" less "Net Fixed Assets." The actual market value of debt (see fn. 7) and equity claims is $74,919 million. The overvalue ratio is 0.98 [= $74,919/($26,640 + $49,779)]. The SIC 26 overvalue ratio is 1.09 [= $40,284/($19,340 + $17,758)] and the SIC 30 overvalue ratio is 0.75 [= $17,000/($8,136 + $14,441)].

TABLE 3—SENSITIVITY OF EQUITY WINDFALLS TO KEY PARAMETERS

	SIC 20	SIC 26	SIC 30
1 Parameters Restricted to Same Value:	None		
Parameters Varying Across Industries:	Beta, Debt Ratio, Rental Price		
Percentage Windfall	−4.88 percent	−6.50 percent	−9.90 percent
2 Parameters Restricted to Same Value:	Beta		
Parameters Varying Across Industries:	Debt Ratio, Rental Price		
Percentage Windfall	−4.88 percent	−6.56 percent	−9.77 percent
3 Parameters Restricted to Same Value:	Beta, Debt Ratio		
Parameters Varying Across Industries:	Rental Price		
Percentage Windfall	−4.88 percent	−6.15 percent	−11.48 percent
4 Parameters Restricted to Same Value:	Beta, Debt Ratio, Rental Price		
Parameters Varying Across Industries:	None		
Percentage Windfall	−4.88 percent	−6.18 percent	−11.98 percent

Notes: When parameters are restricted they are set to the value they have in SIC 20.

A. *Constant Personal Tax and Interest Rates but Full Adjustment to ACRS*

Implementation of the Accelerated Cost Recovery System (ACRS) allowed new investments to be depreciated more quickly than before resulting in an increase in the present value of marginal depreciation tax savings. Old capital is not entitled to the higher tax savings and a wedge is inserted between the values of new and old capital.

The present value of tax depreciation deductions rises from 32 to 47 cents per dollar of structures investment and from 60 to 75 cents for equipment. The response to this incentive effect is a decline in the marginal physical product of capital and an equilibrium where pretax returns may be smaller yet satisfy zero net present value conditions. The smaller rental price and *NOI* translate into a decline in fundamental equity value of \$81.7 billion, a 6.27 percent windfall loss relative to the \$1,303 billion market value of outstanding NFC equity.[9]

The reduction in fundamental equity value represents an "unlevered loss" and occurs regardless of capital structure. With debt financing there is an additional "levered loss." The levered loss can be found by noting that the debt ratio remains unchanged if there is an immediate debt repayment of \$30.8 billion. The retirement reduces future debt services thereby raising discounted equity cash flows by \$19.3 billion. The \$11.5 billion difference is the levered loss, and the proportion $(1-\alpha)/(1-\tau\alpha)$ falls on equity.[10] For the NFC the levered loss is \$9.2 billion, representing 0.71 percent of equity values.

The net impact of ACRS is a \$90.9 billion reduction in fundamental equity value representing 6.98 percent of the market value of outstanding NFC equity (see row 2, Table 2). At the industry level ACRS again inserts a wedge between the values of new and old capital. In SIC 20 the fundamental equity value decline of \$1,858 million represents a 4.88 percent windfall stock market loss. The loss in SIC 26 is \$1,587 million or 6.50 percent of industry stock market value and in SIC 30 the \$532 million loss is 9.90 percent of shareholder value.

The sensitivity of the predicted losses to key parameters is analyzed by reestimating with alternative pre- and post-ACRS parameters. Those results are in Table 3. First, the equity betas in all industries are set to the value of beta in SIC 20. The reduction of beta in SIC 30 by one-fifth lessens the windfall from -9.90 to -9.77 percent (see row 2). In row 3 sensitivity to capital structure is shown by restricting the debt ratio to its value in SIC 20. The decline in debt ratio by one-quarter in SIC 30 increases the equity loss by one-fifth. Row 4 holds pre- and post-ACRS rental prices constant and shows the percentage loss is not sensitive to the level of rental price.

---

[9] David Cutler (1988) examines stock price responses to the Tax Reform Act of 1986. He finds revaluation of share prices from changes in marginal tax depreciation schedules are empirically important.

[10] Given that $\Delta r = \alpha(1-\tau)\Delta i + (1-\alpha)\Delta e$ and the additional risk is shared equally be debt and equity (i.e., $\Delta i = \Delta e$), the proportion of Δr borne by equity is $(1-\alpha)/[\alpha(1-\tau)+(1-\alpha)]$, or $(1-\alpha)/(1-\tau\alpha)$. The proportion borne by debt, $\alpha(1-\tau)/(1-\tau\alpha)$, is an increasing function of the debt ratio.

The impact of ACRS on share prices in Stone, Clay, and Glass is more than in either of the other two industries—5 percentage points more than in Food Products. The overvalue ratio at 0.75 acts as a leverage ratio in that it amplifies the response of security prices to a given change in fundamental value. The large negative windfall occurs in Stone, Clay, and Glass because the market correctly capitalizes the value loss but the undervalued equity base is relatively small.[11]

B. *Constant Interest Rates but Full Adjustment to ACRS and Personal Tax Rates*

The 1981 Tax Act reduced the maximum marginal personal tax rate from 0.70 to 0.50. This exerts an influence on fundamental equity value through τ^i and τ^e. Peek and Wilcox have computed that from 1981 to 1982 the average marginal personal tax rate on interest income falls from 0.366 to 0.333. τ^i is assumed to fall likewise. The computation of τ^e depends on the statutory maximum tax rate (see fn. 3) and is computed to fall from 0.182 to 0.130. There is a resulting fall in equity and average financing rates and rental prices. Row 3 of Table 2 shows there is a further 10-basis point predicted shareholder loss.

C. *Full Adjustment to ACRS, Personal Tax, and Interest Rates*

The ACRS investment incentives shifted the demand schedule for capital and exerted upward pressure on the interest rate. The rising interest rate has two effects. The average financing rate increases and rental prices increase. These two contrary influences tend to reduce fundamental value due to a discounting effect but tend to increase fundamental value due to the *NOI*-rental price effect.

The net impact of a rising interest rate on fundamental equity value may be positive or negative and the direction of influence depends on the interest elasticity (duration) of the marginal return stream relative to the interest elasticity of the average return stream.[12] A stream with a long duration has a higher elasticity and responds more to an interest rate change than does a short duration stream. A rising financing rate reduces the marginal investment's net present value below zero and the producer responds by selecting an investment budget and production plan with an increased marginal product of capital. The duration of the marginal return stream measures the rise in marginal product and rental price that reestablishes zero net present value marginal equilibrium.

The increased rental price increases the total *NOI* stream. This rise offsets a portion of the total value loss brought on by the higher discounting effect. When average duration is less than marginal duration the rise in total fundamental value from a higher rental price is more than the total value loss induced by higher financing rates. A smaller change in rental price—a change proportional to average duration—would have left total value unchanged, but the lesser change would not have reestablished marginal equilibrium.

---

[11] The ratio of ΔV to sales is -0.69, -1.63, and -0.91 percent in SICs 20, 26, and 30, respectively. Thus, the value loss relative to sales in SIC 30 lies between the other industries. The ratio of sales to debt plus equity market value, though, is 3.79, 2.01, and 6.26, respectively. In SIC 30, an unusually small stock of value supports a proportionately large amount of sales.

[12] The average duration of the total capital income stream (equation (5)) is 4.84 years. The duration of the marginal return stream at the zero net present value equilibrium (equation preceding (8)) is 5.62 years. The ratio of average to marginal duration is less than unity and $dV/di > 0$. The relationship between interest rates and duration is discussed by Eugen Bohm von Bawerk (1988, esp. pp. 342–49), wherein it is referred to as "average period." J. R. Hicks (1939, esp. pp. 186–88, 222–24) describes how value and interest rate changes are related to the relative duration of return streams. The computational formula for duration is introduced by Frederick Macaulay (1938).

To deduce the impact of ERTA on interest rates an estimate is borrowed from Patric Hendershott (1986). The Hendershott model allocates a fixed capital stock among thirteen classes of residential and nonresidential capital and finds that a 106-basis point interest rate rise in response to ERTA maintains the aggregate demand for capital at its initial level. Comparison of rows 3 and 4 in Table 2 shows that the impact of a 1.06 percent interest rate rise is a 1.00 percent windfall equity gain in the NFC and in the different industries the gain ranges from 0.67 to 1.23 percent.

The total equity windfall induced by ERTA from ACRS, personal tax, and interest rate changes are $\$-79.1$ billion, a loss equivalent to 6.07 percent of the market value of outstanding equities in the nonfinancial corporate sector. In SICs 20, 26, and 30, the equity windfalls equal -4.27, -5.89, and -8.63 percent of stock market values. Stock returns in two industries are predicted to outperform market; a third industry is predicted to underperform.

IV. A Comparison of Predicted and Actual Stock Price Movements

The ideas in the 1981 tax legislation originated with a bill introduced in 1977. The bill received little attention until President Reagan introduced to Congress on February 18, 1981, a tax proposal adopting some of the earlier provisions. Progress toward tax reform was slow and not until June 1981, did the president and Congress seem serious about compromise. The bill quickly came together and on August 14, President Reagan signed the Economic recovery Tax Act of 1981. This section measures the windfall stock returns accruing because of ERTA between February 18 and August 14, 1981. Our measurement excludes price adjustments prior to February 18 reflecting anticipations about the direction tax policy was taking, or adjustments after August 14 reflecting slow revisions in expectation.

The valuation model can be tested by comparing the predictions with the relative performance of each industry. Separate portfolios are constructed for common stocks in the Food Products (SIC 20), Paper and Allied Products (SIC 26), and Stone, Clay, and Glass (SIC 30) industries. Included securities are listed on the Center for Research in Security Prices (CRSP, 1985) Daily Returns File and in each portfolio there are 79, 42, and 32 securities, respectively. The NFC proxy is the CRSP equal weighted market index.

Standard "event-study" methodologies (see H. Tehranian, N. Travlos, and J. Waegelein, 1987) estimate the excess risk-adjusted return for security j in portfolio n, $ER^{j,n}$, as

$$ER_t^{j,n} = R_t^j - \left(\alpha^j + \beta^j R_t^m\right).$$

R^j and R^m are returns on security j and the market index, respectively, and α^j and β^j are the ordinary least squares estimates of R^j on R^m. Excess returns are estimated for each security in the portfolio for all trading days between February 18 and August 14, 1981.[13]

The cumulative average excess return between days t and u for N securities in industry n, $CAER_{t,u}^n$, is

$$CAER_{t,u}^n = (1/N) \sum_{s=t}^{u} \sum_{j=1}^{N} ER_s^j.$$

$CAER$ measures relative industry windfalls and its expected value is zero if industry returns equal market. The null hypothesis that industry and market perform the same

---

[13] The α and β estimated between August 4, 1980, and January 6, 1981 (106 trading days) is assumed constant until August 14, 1981. As a check for stationarity α and β are reestimated for the period September 8, 1981, to March 1, 1982 (106 trading days). The industry average β in the pre- and post-reform periods are 0.80 and 0.78 in SIC 20, 0.76 and 0.79 in SIC 26, and 0.99 and 0.94 in SIC 30.

TABLE 4—ACTUAL RISK-ADJUSTED EXCESS STOCK RETURNS

Row	SIC 20		SIC 26		SIC 30	
1 Predicted Excess Return						
	1.80 percent		0.18 percent		−2.56 percent	
2 Actual Cumulative Average Excess Return, February 18 to August 14						
	6.71 percent	(3.55)	4.09 percent	(1.30)	−3.82 percent	(−0.62)
3 Actual Cumulative Average Excess Return for Nine ERTA Events						
Total of All Events	2.25 percent	(2.75)	2.06 percent	(1.13)	−2.18 percent	(−1.26)
By Event						
1	0.314 percent	(1.378)	1.287 percent	(2.800)	0.425 percent	(0.354)
2	0.507	(1.665)	−0.141	(0.149)	−0.795	(−1.395)
3	0.263	(1.135)	0.439	(0.841)	0.254	(0.052)
4	0.283	(0.955)	−0.577	(−1.783)	−0.327	(−0.595)
5	−0.247	(−0.858)	0.579	(0.543)	−0.237	(−0.276)
6	0.932	(2.179)	1.242	(2.758)	−0.282	(−0.627)
7	0.415	(1.517)	−0.336	(−0.831)	−0.754	(−0.360)
8	−0.415	(−0.741)	0.083	(0.103)	−0.492	(−0.865)
9	0.203	(1.013)	−0.520	(−1.200)	0.022	(−0.071)

Notes: Predicted excess return equals the difference between the industry and NFC equity windfalls, as listed in row 4, Table 2. Event returns are the cumulative average excess industry stock returns for the day before and day of the event date listed in Table 5. Standard normal statistics testing for zero equality (see equation (13)) are in parentheses. Estimates are constructed from formulas and data described in Section IV.

is tested with the standard normal statistic

$$(13)\quad Z^n = \left[N(u-t+1)\right]^{-1/2} \times \sum_{s=t}^{u} \sum_{j=1}^{N} \left(ER_s^{j,n}/S_s^{j,n}\right).$$

$S^{j,n}$ is the estimated forecast error standard deviation.[14] The null hypothesis that the industry *CAER* equals the prediction from the valuation model is tested with the standard normal statistic

$$Z^p = Z^n(1 - \%\Delta V/CAER^n),$$

where percentΔV is the prediction.

*CAER*s are listed in row 2 of Table 4. SIC 20 outperforms the market by 6.71 percent substantially more than the 1.80 percent prediction listed in row 1. Predicted and actual excess returns in SIC 26 are 0.18 and 4.09 percent, respectively, and in SIC 30 they are −2.56 and −3.82 percent. Performance relative to the market is in all industries the same sign as predicted but the variance of stock returns is large. Only in SIC 20 are excess returns significantly different from zero, as evidenced by the Z^{20} of 3.55 listed in the parentheses of Table 4.

The *CAER*s discussed above include excess returns attributable to non-ERTA information changes. That effect can be dampened by measuring excess returns on only those days when significant news about ERTA becomes available. A review of *The Congressional Quarterly Almanac* (1981) and *The Wall Street Journal Index* (1980, 1981) suggests there are nine releases of significant information. These are listed in Table 5.[15]

---

[14] S^j is computed from the forecast error standard deviation formula given in most econometrics books. This assumes independent and identically normally distributed stock returns. The biasedness introduced by cross-sectional dependence among securities is examined by John Binder (1985) and Paul Malatesta (1986).

[15] The event is the day of publication of announcement in *The Wall Street Journal* but the *CAER* also includes returns on the trading day before the event. Sensitivity is examined by computing *CAER*s for the nine events including three days prior to the event. In SICs 20, 26, and 30 these *CAER*s (and Z statistics) are 3.19 percent (3.07), 1.86 percent (0.99), and −1.64 percent (−0.91), respectively.

TABLE 5—DATES OF IMPORTANT INFORMATION RELEASES ABOUT ERTA

Event	Date	Description
1	2-18-81	Reagan Formally Announces Tax Reform Proposal to Congress
2	6-8-81	Reagan Offers Revised Tax Reform Proposal
3	6-10-81	Compromise Plan on Tax Treatment of Depreciation is Worked Out Between Reagan and Business
4	6-19-81	House Ways & Means Committee Tentatively Approves New Depreciation Guidelines
5	6-25-81	Senate Finance Committee Tentatively Approves New Depreciation Guidelines
6	7-24-81	Reagan and House Ways & Means Committee Work Out a Compromise Bill
7	7-30-81	Senate and House Each Pass Their Own Version of Tax Reform; a Panel Is Set to Resolve Conflicts
8	8-5-81	House and Senate Pass Compromise Bill
9	8-14-81	Reagan Signs Economic Recovery Tax Act of 1981

Notes: The event is the day of announcement publication in *The Wall Street Journal.*

Row 3 of Table 4 presents for each industry the *CAER*s during the nine ERTA "event-windows." In all industries the *CAER*s decline in absolute value relative to the continuous February–August interval, implying excess returns outside the windows were generally the same sign as within. However, within the windows there is less noise as evidenced by the decline in all industries of the implied standard deviation of excess returns (the ratio $CAER/Z$). The lack of statistical significance is still evident in SICs 26 and 30 but in SIC 20 the *CAER* of 2.25 percent is statistically greater than zero ($Z^{20} = 2.75$) but not different from predicted ($Z^{p} = 0.5$).

The valuation model predicts differences between industries. The null hypothesis that the *CAER* in one industry equals the *CAER* in another is tested with the standard normal statistic

$$Z^{n-m}\left[\sum_{s=t}^{u}\left(\sum_{j=1}^{N} ER_s^{j,n}/NS_s^{j,n}-\sum_{j=1}^{M} ER_s^{j,m}/MS_s^{j,m}\right)\right]\Big/[(u-t+1)(1/N+1/M)]^{1/2}.$$

N and M are the number of securities in industries n and m, respectively. The null hypothesis that the difference in *CAER*s between two industries equals the predicted difference is tested with the standard normal statistic

$$Z^{p,n-m}=[Z^{n}(1-\%\Delta V^{n}/CAER^{n})-Z^{m}(1-\%\Delta V^{m}/CAER^{m})]/2^{0.5}.$$

Shareholders in SIC 20 outperform those in SIC 30 by 4.43 percent. That amount is statistically greater than zero ($Z^{20-30} = 2.5$) but indistinguishable from the predicted outperformance of 4.36 percent ($Z^{p,20-30} = 0.2$). Likewise, the actual difference of 4.24 percent between SICs 26 and 30 is significantly different from zero at the 10 percent level ($Z^{26-30} = 1.7$) but not significantly different from predicted ($Z^{p,26-30} = 0.6$). Between SICs 20 and 26 *CAER*s are indistinguishable.

The listings of *CAER*s by event are in Table 4. No single event has a statistically significant impact on all three industries. However, a preponderance of events have a positive impact in SIC 20 and negative impact in SIC 30. The results are not being fixed by a few key events.

The analysis of stock returns data provides moderate support for the predictions from the valuation model. During the continuous February–August interval the performance of each industry relative to the market has the same sign as predicted but they are statistically indistinguishable. When the analysis is restricted to dates of important legislative action the results strengthen. In two of three interindustry comparisons the relative shareholder performance is significantly greater than zero but insignificantly different than predicted.

V. Concluding Remark and Summary

Policymakers affect the allocation of wealth and unanticipated policy changes redistribute it. Sound public policy requires an awareness of who is receiving and who is paying for the free lunch ordered by the policymakers. This study presents and tests a fundamental valuation model for predicting the windfalls resulting from tax policy changes.

The model is applied to estimate fundamental values for the U.S. nonfinancial corporate sector and three manufacturing industries. The ratios of year-end 1980 financial market value to fundamental value equaled 0.81 for the NFC, 0.98 in the Food Products Industry, 1.09 in paper and Allied Products, and 0.75 in Stone, Clay, and Glass.

Fundamental value changed because of the enactment of the Economic Recovery Tax Act of 1981. The largest effect was from the Accelerated Cost Recovery System because new capital received preferential tax treatment and old capital lost value. A slight positive influence on fundamental value was the increasing demand for capital and rising interest rate—new capital was adversely affected and old capital benefited. On net, ERTA favored new capital and the fundamental value of existing capital declined. The stock market at-large is estimated to have suffered a 6.1 percent windfall loss and in the different industries predicted losses range from 1.8 percent less to 2.6 percent more than market.

Portfolios of common stocks were formed in the NFC and in separate industries and actual changes in shareholder wealth were measured. When the analysis is restricted to dates of important legislative action adjustments in shareholder wealth provide moderate support for predictions from the valuation model. In two of three cases interindustry differences in shareholder returns are significantly greater than zero but insignificantly different than predicted. In the third case results are inconclusive. This study suggests that because of the enactment of ERTA shareholders in the Food Products and Paper Products Industries did better than those in Stone, Clay, and Glass, and the relative performance may have been a predictable outcome.

REFERENCES

Auerbach, Alan, J., "Tax Reform and Adjustment Costs: The Impacts on Investment and Market Value," NBER Working Paper No. 2103, December 1986.

______ **and Kotlikoff, Laurence J.,** "Investment versus Savings Incentives: The Size of the Bang for the Buck and the Potential for Self-Financing Business Tax Cuts," in L. H. Meyer, ed., *The Economic Consequences of Government Deficits*, USA: Kluwer-Nijhoff, 1983, 121–49.

Behravesh, N., Hearings before the Subcommittee on Economic Stabilization of the Committee on Banking, Finance, and Urban Affairs, House of Representatives, 99th Congress, Serial No. 99-33, March 13, 1985.

Binder, John J., "Measuring the Effects of Regulations with Stock Price Data," *Rand Journal of Economics*, Summer 1985, *16*, 167–83.

Bohm von Bawerk, Eugen, William Smart, trans., *The Positive Theory of Capital*, New York: G. E. Stechert & Co., 1888.

Brainard, William, Shoven, John and Weiss, Laurence, "The Financial Valuation of the Return to Capital," *Brookings Papers on Economic Activity*, 1980:2, 453–502.

Coen, Robert M., "Investment Behavior: The Measurement of Depreciation, and Tax Policy," *American Economic Review*, March 1975, *65*, 59–74.

Cutler, David M., "Tax Reform and the Stock Market: An Asset Price Approach," *American Economic Review*, December 1988, *78*, 1107–17.

Downs, Thomas W., "Using the User Cost," *Journal of Economics and Business*, November 1986, *38*, 297–305.

______ **and Hendershott, Patric H.**, "Tax Policy and Stock Prices," *National Tax Journal*, June 1987, *40*, 183–90.

Eisner, Robert, "Components of Capital Expenditures: Replacement and Modernization vs. Expansion," *Review of Economics and Statistics*, August 1972, *54*, 297–305.

Feldstein, Martin S. and Rothschild, Michael, "Toward an Economic Theory of Replacement Investment," *Econometrica*, May 1974, *42*, 393–423.

Goulder, Lawrence H. and Summers, Lawrence H., "Tax Policy, Asset Prices, and Growth: A General Equilibrium Analysis," NBER Working Paper No. 2128, July 1986.

Hall, Robert, and Jorgenson, Dale W., "Tax Policy and Investment Behavior," *American Economic Review*, June 1967, *57*, 391–414.

Hendershott, Patric H., "Tax Changes and Capital Allocation in the 1980s," in M. Feldstein, ed., *The Effect of Taxation on Capital Formation*, Chicago: University of Chicago Press, 1986.

Hicks, John R., *Value and Capital*, London: Oxford University Press, 1939.

Holland, Daniel M. and Myers, Stewart C., "Profitability and Capital Costs for Manufacturing Corporations and All Nonfinancial Corporations," unpublished manuscript, MIT 1981.

Hulten, Charles R. and Wykoff, Frank C., "The Measurement of Economic Depreciation," in C. Hulten, ed., *Depreciation, Inflation, and the Taxation of Income from Capital*, Washington: Urban Institute Press, 1981, 81–125.

Ibbotson, Roger G., *Stocks, Bonds, Bills, and Inflation: 1984 Yearbook*, Chicago: R. G. Ibbotson Associates, 1984.

Macaulay, Frederick R., *Some Theoretical Problems Suggested by the Movement of Interest Rates, Bond Yields, and Stock Prices in the United States Since 1856*, New York: Columbia University Press for the NBER, 1938.

Malatesta, Paul, "Measuring Abnormal Performance: The Event Parameter Approach Using Joint Generalized Least Squares," *Journal of Financial and Quantitative Analysis*, March 1986, *21*, 27–38.

Modigliani, Franco and Miller, Merton, "Corporate Income Taxes and the Cost of Capital: A Correction," *American Economic Review*, June 1963, *53*, 433–43.

Peek, Joe and Wilcox, James A., "The Postwar Stability of the Fisher Effect," *Journal of Finance*, September 1983, *38*, 1111–24.

Tehranian, Hassan, Travlos, Nickolaos and Waegelein, James, "The Effect of Long-Term Performance Plans on Corporate Sell-Off-Induced Abnormal Returns," *Journal of Finance*, September 1987, *42*, 933–42.

VonFurstenberg, George M., Malkiel, Burton G. and Watson, Harry S., "The Distribution of Investment between Industries," in G. M. VonFurstenberg, ed., *Capital, Efficiency, and Growth*, Cambridge, MA: Ballinger, 1980.

Center for Research in Security Prices, Daily Stock Returns File, University of Chicago, 1985.

Congressional Quarterly Almanac, *Congressional Quarterly, Inc., Washington, D.C.*, 1981.

Moody's Investors Service, *Moody's Bond Record*, New York, NY, July 1981.

Standard & Poor's *COMPUSTAT* Services, Expanded Industrial Annual File, New York, NY, 1984.

U.S. Department of Commerce, Bureau of Economic Analysis, "Fixed Nonresidential Capital Stock Estimates," 1986.

U.S. Federal Reserve Board, SSRC-MIT-PENN Quarterly Econometric Model, Washington: USGPO, September 1983.

U.S. Federal Reserve Board of Governors, *Balance Sheets for the U.S. Economy, 1945–84*, Flow of Funds Section, October 1985.

Wall Street Journal Index, Dow Jones and Company, New York, 1980, 1981.

Third-Degree Price Discrimination, Welfare and Profits: A Geometrical Analysis

By STEPHEN LAYSON*

This note introduces a new geometrical method of measuring the welfare effects of third-degree price discrimination that builds on earlier work by Richard Schmalensee (1981). The geometrical technique presented here suggests a simple empirically useful formula for the calculation of the welfare effects of third-degree price discrimination under linear demand. This geometrical technique also demonstrates the special relationship between the welfare and profit effects of price discrimination under linear demand and in this regard corrects an error in Hal Varian (1985).

Figure 1 illustrates the profit-maximizing solution for a monopolist that faces two independent linear demands. In Figure 1 the demand and marginal revenue curves for group 1 are drawn with respect to the left-hand side origin, O_1, while the demand and marginal revenue curves for group 2 are drawn with respect to the right-hand side origin, O_2. The width of Figure 1 represents the total profit-maximizing output under both price discrimination and simple monopoly.[1] At this output level marginal cost is MC_0.

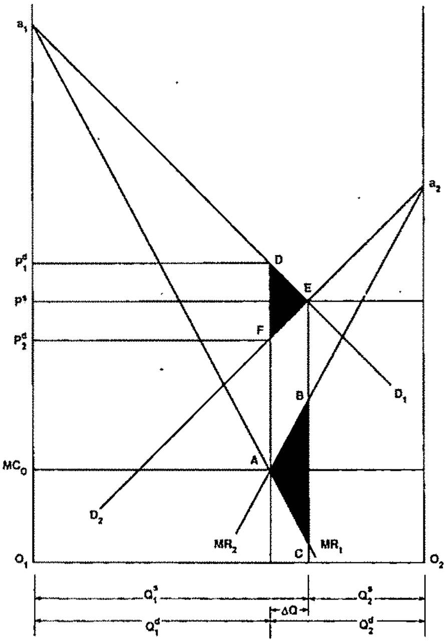


FIGURE 1

Under simple monopoly each group is charged a price per unit of P^s. At this price group 1 purchases Q_1^s and group 2 purchases Q_2^s. In general the marginal revenues of selling to the two groups will not be the same at the simple monopoly allocation. For example, in Figure 1, $MR_2 > MR_1$ at the simple monopoly allocation. Starting from this situation the price-discriminating monopolist will increase sales to group 2 to Q_2^d and decrease sales to group 1 to Q_1^d. At this allocation $MR_1 = MR_2 = MC_0$ and profits are maximized under price discrimination. The prices charged to groups 1 and 2 under price discrimination, are, respectively, P_1^d and P_2^d.

Because total output is the same under price discrimination and simple monopoly, the profit gain from price discrimination is the additional revenue from increased sales to group 2 minus the revenue foregone from reduced sales to group 1. In Figure 1 the shaded triangular area *ABC* represents the profit gain from price discrimination. The welfare loss from price discrimination in Figure 1 is represented by the shaded triangular area between the two demand curves

---

*Department of Economics, University of North Carolina at Greensboro, Greensboro, NC 27412. I wish to thank John P. Formby for stimulating my interest in this subject and for helpful comments.

[1] See Joan Robinson (1933, p. 192) for a proof of this proposition. It is assumed in this note that both markets are served under simple monopoly.

labeled *DEF*. This welfare loss stems from the transfer of ΔQ units from group 1, who values the good more highly, to group 2 who values the good less highly.[2]

Referring to Figure 1, the welfare loss triangle and the profit triangle have heights of equal length, ΔQ. The base of the profit triangle, however, is twice the base of the welfare loss triangle. The reason for this is that the slopes of the marginal revenue curves are twice the slopes of their respective demand curves. It follows that the area of the profit triangle is exactly twice the area of the welfare loss triangle. Varian's (1985, p. 875) result that the welfare loss is equal to the profit gain in the linear demand case is incorrect.[3]

The mathematical derivations of the profit gain and welfare loss from price discrimination in the linear demand case are straightforward and revealing. Let the demand curves be given by

$$(1)\quad P_i = a_i - b_i Q_i; \quad a_i, b_i > 0 \; i = 1,2.$$

Assuming profit maximization and linear demands, it can be shown that the relationship between the profit change, $\Delta\pi$, and the welfare change, ΔW, from price discrimination is given by[4]

$$(2)\quad \Delta W = -1/2\,\Delta\pi$$

$$= -(a_1 - a_2)^2/[(8(b_1 + b_2)].$$

The profit gain and the welfare loss are directly proportional to the squared difference in the price intercepts of the demand curves and inversely proportional to the sum of the demand slopes. For the linear demand case we see from equation (2) that price discrimination is most harmful when it is most profitable, and hence most likely to be undertaken. An alternative empirically useful formula for the welfare change suggested by the triangular measure in Figure 1 is $-1/2(P_1^d - P_2^d)\Delta Q$. Solving for ΔQ in terms of observable variables under price discrimination yields equation (3):

$$(3)\quad \Delta W = -\tfrac{1}{2}Q_1^d Q_2^d \left(P_1^d - P_2^d\right)^2 /$$

$$\left[Q_1^d\left(P_2^d - MC_0\right) + Q_2^d\left(P_1^d - MC_0\right)\right].$$

Note that calculation of the expression for the welfare change in equation (3) does not require that one observe the transition from simple monopoly to price discrimination. Data are required only on variables observed under price discrimination. Furthermore, equation (3) does not require that the demand curves be globally linear; it requires linearity only in the output range between simple monopoly and price discrimination. The validity of equation (3) like equation (2) discussed previously assumes profit maximization, however.

The geometrical technique presented in this note may also be used to measure the welfare effects of price discrimination when demands are nonlinear. When demands are nonlinear there are two welfare effects: (1) a welfare effect due to the transfer of sales from one group to the other, and (2) a welfare effect due to the net change in total output. For a discussion of the nonlinear case and other extensions, see my paper (1987).

REFERENCES

Layson, Stephen, "Third Degree Price Discrimination Welfare and Profits: A Geometrical Analysis," Working Paper Series No. ECO870402, Center for Applied Research, University of North Carolina-Greensboro, April 1987.

Robinson, Joan, *The Economics of Imperfect Competition*, London: Macmillan, 1933.

Schmalensee, Richard, "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review*, March 1981, *71*, 242–47.

Varian, Hal, "Price Discrimination and Social Welfare," *American Economic Review*, September 1985, *75*, 870–75.

---

[2] The interested reader may wish to compare Figure 1 in this note to the geometric treatment in Schmalensee (1981, p. 246).

[3] His mathematical demonstration shows only that $\Delta W \geq -\Delta\pi$ for the linear demand case.

[4] Derivations of equations (2) and (3) are available from the author on request.

Cooperative and Noncooperative *R&D* in Duopoly with Spillovers

By CLAUDE D'ASPREMONT AND ALEXIS JACQUEMIN*

Contrary to the usual assumption made in most oligopoly models, relations among firms are seldom of a wholly cooperative or noncooperative type: in many situations, they compete in some fields, while they cooperate in others. An important example is the case of cooperative research efforts bringing fierce competitors together.

Two types of agreement are observed. First *R&D* cooperation can take place at the so-called "precompetitive stage": companies share basic information and efforts in the *R&D* stage but remain rivals in the marketplace.[1]

A second type of agreement involves an extended collusion between partners, creating common policies at the product level. The usual justifications of this extension are the difficulties of protecting intellectual property. The idea is then to allow partners who have achieved inventions together, to also control together the processes and products which embody the results of their collaboration, in order to recuperate jointly their *R&D* investments.[2]

What could be expected from these types of agreement is a reduction in *R&D* expenditures, because of less wasteful duplication, and a reduction of total production, because of more monopoly power. Using a two-stage approach, this paper provides an example that does not fulfill these expectations and that allows a social welfare comparison between the corresponding games. An important factor in this analysis consists in the externalities or spillovers in *R&D* from one firm to another.

I. The Example

Consider an industry with two firms facing an inverse demand function $D^{-1}(Q)$, where $Q = q_1 + q_2$ is the total quantity produced. Each firm has a cost of production $C_i(q_i, x_i, x_j)$ which is a function of its own production, q_i, of the amount of research x_i that it undertakes and the amount of research x_j that its rival undertakes. Both D^{-1} and C are assumed linear, so that

$$D^{-1} = a - bQ \quad \text{with } a, b > 0, \quad \text{and}$$

$$C_i(q_i, x_i, x_j) = \left[A - x_i - \beta x_j\right] q_i, \qquad i = 1,2,\ i \neq j$$

with $0 < A < a$, $0 < \beta < 1$; $x_i + \beta x_j \leq A$; $Q \leq a/b$.

The *R&D* externalities or spillovers imply that some benefits of each firm's *R&D* flow without payment to other firms. In our specification the external effect of firm j *R&D* is to lower firm i's unit production cost.[3] The cost of *R&D* is assumed to be

---

*Center for Operations Research & Econometrics, 1348 Louvain-La-Neuve, Belgium. We are grateful to Jean Gabszewicz for his comments. Also the first author acknowledges support from the CIM, Belgium, and the hospitality of the Graduate School of Business, Stanford University.

[1] A well-known example is the European Strategic Program for *R&D* in Information Technologies (ESPRIT). The Microelectronics and Computer Technology Corporation (MCC) in the United States, and the Very Large-Scale Integration Program (VLSI) in Japan are other examples.

[2] An illustration is given by the EUREKA project intended to undertake common European research with concrete objectives and leading to the joint exploitation of the results.

[3] One interpretation is that successful inventions of rivals can be imitated at less cost to firm i than if it were to invent the new processes itself. See J. Hartwick (1984) who presents a static Cournot model with free entry and concludes that with *R&D* spillovers one cannot conclude that the corresponding equilibrium is associated with excessive duplication as Dasgupta and Stiglitz (1980).

quadratic, reflecting the existence of diminishing returns to *R&D* expenditures.[4]

Firms' strategies consist of a level of research and a subsequent production strategy based on their *R&D* choice. We shall now analyze three different games.

1. In the first one, firms act noncooperatively in both output and *R&D*. Consider the profit of firm i at the second stage, conditional on x_1 and x_2:

$$\pi_i = [a - bQ]q_i - [A - x_i - \beta x_j]q_i - \gamma\frac{x_i^2}{2}, \qquad j \neq i,\ i = 1,2.$$

The Nash-Cournot equilibrium can be computed to be[5]

$$q_i = \frac{(a-A)+(2-\beta)x_i+(2\beta-1)x_j}{3b}.$$

At the preceding stage, in which firms choose *R&D* levels, profits can be written as

$$\pi_i^* = \frac{1}{9b}\left[(a-A)+(2-\beta)x_i+(2\beta-1)x_j\right]^2 - \gamma\frac{x_i^2}{2} \qquad j \neq i,\ i = 1,2.$$

This integrates a triple influence of the *R&D* levels: via the outputs, the unit production cost, and the *R&D* costs themselves.

There exists a unique (and symmetric) solution[6] satisfying $\partial\pi_i^*/\partial x_i = 0$, for which

$$x_i^* = \frac{(a-A)(2-\beta)}{4.5b\gamma-(2-\beta)(1+\beta)} \qquad i = 1,2.$$

$$Q^* = q_i^* + q_j^* = \frac{2(a-A)}{3b} + \frac{2(\beta+1)}{3b}x_i^* = \frac{2(a-A)}{3b}\left[\frac{4.5b\gamma}{4.5b\gamma-(2-\beta)(1+\beta)}\right].$$

2. In the second game, we introduce cooperation in *R&D*, the second stage remaining noncooperative. At the first stage the firms maximize the joint profits, as a function of x_1 and x_2:

$$\hat{\pi} = \pi_1^* + \pi_2^* = \frac{1}{9b}\sum_{i=1}^{2}\left\{\left[(a-A)+(2-\beta)x_i + (2\beta-1)x_j\right]^2 - \gamma\frac{x_i^2}{2}\right\}, \qquad j \neq i.$$

Considering the symmetric solution $x_1 = x_2 = \hat{x}$, we obtain[7] the following unique solution for the equilibrium with cooperation in *R&D*:

$$\hat{x} = \frac{(\beta+1)(a-A)}{4.5b\gamma-(\beta+1)^2};$$

$$\hat{Q} = \frac{2(a-A)}{3b} + \frac{2(\beta+1)}{3b}\cdot\hat{x} = \frac{2(a-A)}{3b}\left[\frac{4.5b\gamma}{4.5b\gamma-(1+\beta)^2}\right].$$

---

[4] A valuable justification of this assumption is that "the technological possibilities linking *R&D* inputs and innovative outputs do not display any economies of scale with respect to the size of the firm in which *R&D* is undertaken" (Dasgupta, 1986, p. 523).

[5] Notice that

$$q_1 + q_2 \le \frac{1}{3b}[2(a-A)+2A] \le \frac{a}{b}.$$

[6] Second-order conditions require that

$$\frac{2(2-\beta)^2}{9b} - \gamma < 0$$

or $\frac{2}{9}(2-\beta)^2 < b\gamma$.

[7] Second-order conditions require $\frac{2}{9}(1+\beta)^2 < b\gamma$.

These solutions correspond to an internalization of the *R&D* external effects through joint decision on the levels of *R&D* expenditures. Contrary to what could have been expected from possible reduction in the duplication of *R&D*, especially in the case of large spillovers, a comparison between $\hat{x}$ and x^* clearly indicates that for large spillovers, that is,[8] $\beta > 0.5$, the level of *R&D* increases when firms cooperate in *R&D*, that is, $\hat{x} > x^*$. In the same perspective, as shown by the respective values of $\hat{Q}$ and Q^*, the amount of production is also higher with cooperation in *R&D*, than in the noncooperative situation, that is, $\hat{Q} > Q^*$.

To the extent that profits are higher in the case of cooperative research than in the noncooperative game[9] private incentives, independently of any public policy such as subsidies, can be sufficient to lead to such a cooperation.

3. The third case deals with monopoly: Firms cooperate in both stages of the game. At the second stage, the joint profit conditional on x_1 and x_2 is given by

$$\pi = [a - bQ]Q - AQ$$
$$+ (x_1 + \beta x_2)q_1 + (x_2 + \beta x_1)q_2 - \gamma \sum_{i=1}^{2} \frac{x_i^2}{2}.$$

For $x_2 = x_1 = x$, the symmetric solution $\tilde{q}_1 = \tilde{q}_2$ leads to[10]

$$Q = q_1 + q_2 = [(a - A) + (1+\beta)x]/2b.$$

At the preceding stage, the joint profit becomes

$$\tilde{\pi} = \frac{1}{b}\left[\frac{a - A + (1+\beta)x}{2}\right]^2 - \gamma x^2.$$

The symmetric cooperative equilibrium in *R&D* and in production corresponds to the following unique solution[11]

$$\tilde{x} = \frac{(a-A)(1+\beta)}{4b\gamma - (1+\beta)^2};$$

$$\tilde{Q} = \frac{(a-A)}{2b} + \frac{(1+\beta)}{2b}\cdot\tilde{x}$$
$$= \frac{a-A}{2b}\left[\frac{4b\gamma}{4b\gamma - (1+\beta)^2}\right].$$

Here as expected, the collusive output, for a given level of *R&D*, is smaller than the noncooperative one, but it is not necessarily so when the optimal amount of *R&D* is incorporated.[12] Similarly, the collusive amount of *R&D* varies with the value of β and is, for reasonably large spillovers, higher than in the fully noncooperative equilibrium.[13] Furthermore the amount of *R&D* in the case of collusion in both output and *R&D* is higher than in the case of pure *R&D* cooperation. This stems from the fact that less competition in the product market allows the firms to capture more of the surplus created by their research and induce more *R&D* expenditures. But despite this larger amount of *R&D*, the quantity produced at the fully

---

[8] Indeed, if $\beta + 1 > 2 - \beta$, $x^* < \hat{x}$.

[9] Indeed the noncooperative choice could always be adopted by cooperating firms, if more profitable.

[10] Notice that

$$\tilde{Q} \le \frac{1}{2b}[2(a-A) + 2A] = \frac{a}{b}.$$

[11] According to the second-order conditions, $\partial^2\tilde{\pi}/\partial\tilde{x}^2 = (1+\beta)^2/2 - 2\gamma < 0$ or $(1+\beta)^2/4 < b\gamma$. Also we consider here only the *symmetric* cooperative equilibrium. The producers could reach higher joint profits by having different *R&D* expenditures and only one firm producing (the one with the lower unit cost). However, to consider this asymmetric cooperative solution would not affect our qualitative results.

[12] We have $\tilde{Q} < Q^*$ iff $5\beta^2 + 4\beta - 1 < 3b\gamma$. For $\beta = 1$, $\tilde{Q} < Q^*$ would require that $b\gamma > \frac{8}{3}$, which is more restrictive than second-order conditions.

[13] It appears that $\tilde{x} < x^*$

$$\text{iff } \frac{(1+\beta)}{4b\gamma - (1+\beta)^2} > \frac{(2+\beta)}{4.5b\gamma - (2-\beta)(1+\beta)^2} \quad \text{or}$$

$$\beta > 0.41.$$

cooperative equilibrium is less than with a cooperation limited to the *R&D* stage.[14]

II. Welfare Conclusions

Given this set of results, it is not a priori clear that, from a social welfare point of view, one type of behavior is more efficient than another. Indeed more cooperation could lead to higher profits but lower consumer surplus. Less production could be compensated by more *R&D*. And a higher level of research could correspond to a wasteful duplication that ignores *R&D* externalities. In order to classify the solution obtained in the different situations, we need an efficiency standard. Let us define social welfare $W(Q)$ as the sum of the consumer's surplus $V(Q)$ and the producer's surplus (assuming $x_1 = x_2 = x$)

$$W(Q) = V(Q) - AQ + (1+\beta)xQ - \gamma x^2.$$

Given x, the efficient output is the following

$$Q = \frac{1}{b}[a - A + (1+\beta)x].$$

At the first stage social welfare is

$$W^{**} = V(Q) - AQ + (1+\beta)xQ - \gamma x^2.$$

The efficient level of *R&D* satisfying the first-order conditions[15] is

$$x^{**} = \frac{(a-A)(1+\beta)}{2b\gamma - (\beta+1)^2}.$$

And finally the socially efficient amount of production incorporating the efficient level of research can be written as

$$Q^{**} = \frac{a-A}{b} + \frac{1+\beta}{b} \cdot x^{**}$$

$$= \frac{a-A}{b}\left[\frac{2b\gamma}{2b\gamma - (1+\beta)^2}\right].$$

Therefore the solution obtained by maximizing social welfare requires not only more production but also a higher level of *R&D* than what is obtained with any of the previous noncooperative and cooperative equilibria.

Indeed, $x^{**} > x^*$, since

$$\frac{1+\beta}{2b\gamma - (1+\beta)^2} > \frac{2-\beta}{4.5b\gamma - (2-\beta)(1+\beta)}$$

and $$Q^{**} > Q^*.$$

Similarly $x^{**} > \tilde{x} > \hat{x}$ and $Q^{**} > \hat{Q} > \tilde{Q}$.

This provides us with a convenient social efficiency standard to classify our various results.

The clearest conclusion is that cooperation in *R&D* (but not in production) increases both expenditures in *R&D* and quantities of production, with respect to the noncooperative solution, that is, $\hat{x} > x^*$ and $\hat{Q} > Q^*$, whenever the spillover effect is large enough; otherwise it is the reverse.

Further, considering separately production and *R&D* aspects, the cases of large and small spillovers should also be distinguished. For large spillovers, such that $\beta > 0.5$, the amount of research which is the closest to the social optimum is the one achieved by firms' cooperating in both output and research, and the most distant, the one obtained by noncooperative behavior. The complete classification is the following

$$x^{**} > \tilde{x} > \hat{x} > x^*.$$

Concerning the quantity of production, the closest to the social optimum is what is produced by firms' cooperating at the "precompetitive stage," that is, in research. The classification is then

---

[14] Indeed $\tilde{Q} < \hat{Q}$ whenever $b\gamma > (1+\beta)^2/3$, a less restrictive condition than some of our second-order conditions.

[15] The second-order condition requires that $[(1+\beta)/2]^2 < b\gamma$.

$$Q^{**} > \hat{Q} > Q^{*} > \tilde{Q}.$$

For small spillovers, such that $\beta \leq 0.4$, the classifications are different but the "second-best" for *R&D* is still obtained by a cooperative behavior in both stages.[16]

To conclude our example has shown that cooperative behavior can play a positive role in industries having a few firms and characterized by *R&D* activities generating spillover effects. This is in line with the permissive American and European antitrust regulations allowing cooperative research whereby member firms agree to share the costs and the results of a research project.[17]

However, in order to compute explicitly and to classify our various types of subgame-perfect solutions, our analysis has been very partial and based on a model ignoring many crucial aspects of *R&D* activities.[18]

REFERENCES

Davies, S., *The Diffusion of Process Innovations*, Oxford: Cambridge University Press, 1979.

Dasgupta, P. and Stiglitz, J., "Industrial Structure and the Nature of Innovative Activity," *Economic Journal*, June 1980, *90*, 266–93.

______, "The Theory of Technological Competition," in J. Stiglitz and F. Mathewson, eds., *New Developments in the Analysis of Market Structure*, Cambridge: MIT Press, 1986.

De Bondt, R., Sleuwaegen, L. and Veugelers, R., "Innovative Strategic Groups in Multinational Industries," *European Economic Review*, April 1988, *32*, 905–26.

Hartwick, J., "Optimal *R&D* Levels When Firm j Benefits from Firm i's Inventive Activity," *Economic Letters*, 1984, *16*, 165–70.

Jacquemin, A., *The New Industrial Organization*, Cambridge: MIT Press, 1987, ch. 6.

______, "Cooperative Agreements in *R&D* and European Antitrust Policy," *European Economic Review*, March 1988, *32*, 551–60.

Katz, M., "An Analysis of Cooperative Research and Development," *Rand Journal of Economics*, Winter 1986, *17*, 527–43.

Spence, M., "Cost Reduction, Competition and Industry Performance," in J. Stiglitz and F. Mathewson, eds., *New Developments in the Analysis of Market Structure*, Cambridge, MIT Press, 1986.

---

[16] The classifications are then $x^{**} > \tilde{x} \geq x^{*} > \hat{x}$; $Q^{**} > Q^{*} > \hat{Q} > \tilde{Q}$.

[17] Policy questions are analyzed in A. Jacquemin (1987, 1988).

[18] For a broader analysis, see M. Katz (1986). An overview of the main aspects of *R&D* activities and innovations is to be found in S. Davies (1979) and M. Spence (1986). Multinational aspects are considered in R. De Bondt et al. (1988).

Productivity Growth, Convergence, and Welfare: Comment

By J. BRADFORD DE LONG*

Economists have always expected the "convergence" of national productivity levels. The theoretical logic behind this belief is powerful. The per capita income edge of the West is based on its application of the storehouse of industrial and administrative technology of the Industrial Revolution. This storehouse is open: modern technology is a public good. The benefits of tapping this storehouse are great, and so nations will strain every nerve to assimilate modern technology and their incomes will converge to those of industrial nations.

William Baumol (1986) argues that convergence has shown itself strongly in the growth of industrial nations since 1870.[1] According to Baumol, those nations positioned to industrialize are much closer together in productivity now than a century ago. He bases this conclusion on a regression of growth since 1870 on 1870 productivity for sixteen countries covered by Angus Maddison (1982).[2]

Baumol's finding of convergence might—even though Baumol himself does not believe that it should—naturally be read to support two further conclusions. First, slow relative growth in the United States since World War II was inevitable: convergence implies that in the long run divergent national cultures, institutions, or policies cannot sustain significant productivity edges over the rest of the developed world.[3] Second, one can be optimistic about future development. Maddison's sixteen all assimilated modern technology and converged: perhaps all developing nations will converge to Western living standards once they acquire a foundation of technological literacy.[4]

But when properly interpreted Baumol's finding is less informative than one might think. For Baumol's regression uses an *ex post* sample of countries that are now rich and have successfully developed. By Maddison's choice, those nations that have not converged are excluded from his sample because of their resulting present relative

---

*Harvard University, Cambridge, MA 02138, and National Bureau of Economic Research. I would like to thank Moses Abramovitz, Paul David, Christina Romer, David Romer, Paul Romer, Peter Temin, Robert Waldmann, Jeffrey Williamson, and especially William Baumol for helpful comments.

[1] Consider Baumol (1986): "Among the main observations...is the remarkable convergence....[T]here is a strong inverse correlation between a country's productivity...in 1870 and its...productivity growth since then," and Baumol (1987): "Even more remarkable...is the convergence in...living standards of the leading industrial countries.... In 1870...productivity in Australia, the leader, was 8 times...Japan's (the laggard). By 1979, the ratio...had fallen to about two.'

[2] Moses Abramovitz (1986) follows the behavior of these sixteen over time and notes that even among these nations "convergence" is almost entirely a post-World War II phenomenon. Abramovitz' remarks on how the absence of the "social capability" to grasp the benefits of the Industrial Revolution may prevent even nations that could benefit greatly from industrializing are well worth reading. Also very good on the possible determinants of the social capability to assimilate technology are Irma Adelman and Cynthia Taft Morris (1980), Gregory Clark (1987), and Richard Easterlin (1981).

[3] Baumol (1987): "America's [post World World II] lag in productivity growth...is very real.... But we can see this in new perspective as a necessary consequence of convergence." Baumol (1986): "The convergence of productivity levels in industrial nations inevitably condemned those with high 1870 productivity levels to relatively slow productivity growth since then."

[4] Nineteenth century economists like John Stuart Mill [1848] (1970) and Karl Marx [1853] (1973) were extremely optimistic about future convergence. Twentieth century economic historians like Alexander Gerschenkron (1962) stress the extremely rapid growth possible for a latecomer to development, which can accomplish in one generation what took earlier developers four. Baumol (1986) is not so optimistic and so does not draw out this second implication. After examining post-World War II data, he writes that "the poorer less developed countries are still largely barred from the homogenization process.... [P]art of the explanation may well be related to produce mix and education. A less developed country that produces no cars cannot benefit from the invention...of a better car producing robot...[or] from the factor-price equalization effects of the accompanying investments."

TABLE 1—REGRESSIONS USING MADDISON'S SIXTEEN

Independent Variable	Dependent Variable	Constant	Slope Coefficient	Standard Error of Estimate	R^2
Natural Log of 1870 Productivity	Annual Percent Productivity Growth	5.251	−0.749 .075	.14	.87
Natural Log of 1870 Income	Log Difference of 1979 and 1870 Income	8.457	−0.995 .094	.15	.88

Source: Data from Maddison (1982).

poverty. Convergence is thus all but guaranteed in Baumol's regression, which tells us little about the strength of the forces making for convergence among nations that in 1870 belonged to what Baumol calls the "convergence club."

Only a regression run on an *ex ante* sample, a sample not of nations that have converged but of nations that seemed in 1870 likely to converge, can tell us whether growth since 1870 exhibits "convergence." The answer to this *ex ante* question—have those nations that a century ago appeared well placed to appropriate and utilize industrial technology converged?—is no. An unbiased sample of nations relatively rich, well-integrated into the world economy, and thus well-positioned to utilize modern technology as of 1870 have not converged.[5]

Maddison (1982) compiles long-run national income and aggregate productivity data for sixteen successful capitalist nations.[6] Because he focuses on nations which (a) have a rich data base for the construction of historical national accounts and (b) have successfully developed, the nations in Maddison's sixteen are among the richest nations in the world today. Baumol regresses the average rate of annual labor productivity growth over 1870–1979 on a constant and on the log of labor productivity in 1870 for this sample. He finds the inverse relationship of the first line of Table 1. The slope is large enough to erase by 1979 almost all initial income gaps, and the residual variance is small.

Regressing the log difference in per capita income between 1870 and 1979 on a constant and the log of per capita income in 1870 provides a slightly stronger case for convergence, as detailed in the second line of Table 1 and in Figure 1. The logarithmic income specification offers two advantages. The slope has the intuitive interpretation that a value of minus one means that 1979 and 1870 relative incomes are uncorrelated, and extension of the sample to include additional nations becomes easier.

Baumol's regression line tells us little about the strength of forces making for convergence since 1870 among industrial nations. The sample suffers from selection bias, and the independent variable is unavoidably measured with error. Both of these create the appearance of convergence whether or not it exists in reality. Sample selection bias arises because any nations relatively rich in 1870 that have not converged fail to make it into Maddison's sixteen. Maddison's sixteen thus include Norway but not Spain, Canada but not Argentina, and Italy but not Ireland.

A fair test of convergence requires not an *ex post* sample of countries that have converged but an *ex ante* sample of countries that in 1870 looked likely to converge. Moreover, least squares is not a satisfactory estimation technique because of errors in measuring 1870 incomes. Such errors induce

---

[5] Baumol accepts the validity of this basic criticism of Baumol (1986). Baumol and Edward Wolff (1988) and Wolff (1988) show that one can find a small number *n* so that an *ex ante* sample of the richest *n* nations does exhibit convergence. More striking is their documentation of strong convergence among European nations since 1950. Poor European nations have grown much faster than middle-income Latin American nations.

[6] Maddison's focus on nations that have been economically successful is deliberate; his aim in (1964), (1982), and (1987) is to investigate the features of successful capitalist development. In works like Maddison (1970, 1983) he has analyzed the long-run growth and development of less successful nations.

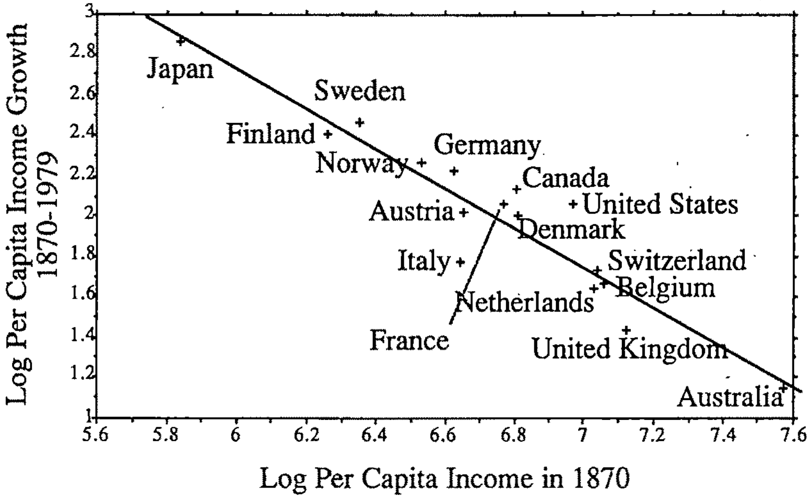


FIGURE 1. PER CAPITA GNP REGRESSION FOR MADDISON'S SIXTEEN

opposite errors in 1870–1979 growth and bias the regression slope toward −1. As Baumol notes, such errors can produce the illusion of an inverse relationship between income in 1870 and growth since.

The unbiased sample used here meets three criteria. First, it is made up of nations that had high potential for economic growth as of 1870, in which modern economic growth had begun to take hold by the middle of the nineteenth century. Second, inclusion in the sample is not conditional on subsequent rapid growth. Third, the sample matches Baumol's as closely as possible, both because the best data exist for Maddison's sixteen and because analyzing an unbiased sample close to Baumol's shows that different conclusions arise not from different estimates but from removing sample selection and errors in variables' biases.

Per capita income in 1870 is an obvious measure of whether a nation was sufficiently technologically literate and integrated into world trade in 1870 to be counted among the potential convergers. Nations with high incomes in 1870 were nations with the material and human resources to industrialize. Modern economic growth had already pushed real incomes far above the levels of the preindustrial world. And such a sample does not exclude nations which had good industrialization prospects in 1870 that have not since fulfilled their potential.[7]

The construction of this sample requires judgment. Per capita income in 1870 must be estimated for nations in the extended sample but not in Maddison's sixteen. The estimation of 1870 income is discussed in the Appendix.[8] Changes in national boundaries must be dealt with; this paper uses modern boundaries throughout. The level of 1870 income to serve as a cutoff for inclusion in the sample must be set. The choice of cutoff level itself requires balancing three goals: including only nations which really did in

---

[7] Alternative measures of prospects for development in 1870, such as per capita industrial production or the proportion of the labor force in agriculture, would serve as well but would make little quantitative difference. The correlations for the sample of Maddison's sixteen between 1870 per capita GNP and 1870 labor productivity and share of the labor force in agriculture are .98 and .84, respectively.

[8] The estimates of 1870 per capita income arrived at in the Appendix are not precise enough to be used for assessing the history and development of any individual country. They do, however, serve adequately as the raw material for a comparative exercise like that carried out here in which explicit econometric correction is made for errors in variables and in which errors in measuring nineteenth century per capita income for any one nation can have only a limited effect.

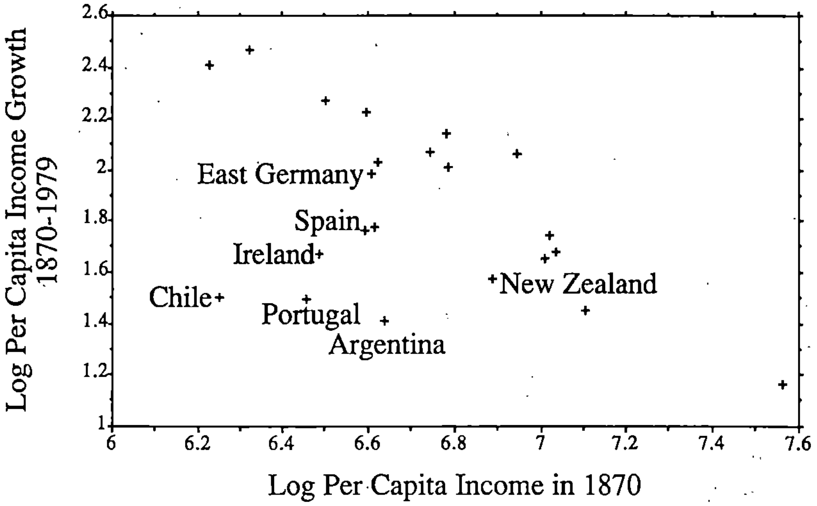


FIGURE 2. 1870 PER CAPITA INCOME AND SUBSEQUENT GROWTH FOR THE ONCE-RICH TWENTY-TWO

1870 possess the social capability for rapid industrialization; including as many nations in Baumol's sample as possible; and building as large a sample as possible.

One cannot proceed by pursuing this last goal at the expense of the others: one should not form a regression sample by including all nations for which 1870 income estimates can be generated. Few would argue that the failure of, say, India to converge is evidence against the convergence hypothesis. Even if nations that were not seen as having high growth potential are removed, an all-inclusive sample suffers from selection bias. Long-run national accounts are luxuries. Nations likely to have the historians and archives necessary to construct such accounts are nations that have converged.

If the convergence club membership cutoff is set low enough to include all Maddison's sixteen, then nations with 1870 incomes above 300 1975 dollars are included. This sample covers half the world. All Europe including Russia, all of South America, and perhaps others (Mexico and Cuba?) were richer than Japan in 1870. This sample does not provide a fair test of convergence. The Japanese miracle is a miracle largely because there was little sign in 1870 that Japan—or any nation as poor as Japan—was a candidate for rapid industrialization.

The second poorest of Maddison's sixteen in 1870 was Finland. Taking Finland's 1870 income as a cutoff leads to a sample in which Japan is removed, while Argentina, Chile, East Germany,[9] Ireland, New Zealand, Portugal, and Spain are added. Growth and initial per capita income levels for this resulting "once-rich twenty-two" sample are plotted in Figure 2.[10]

---

[9] Perhaps only nations that have remained capitalist should be included in the sample, for occupation by the Red Army and subsequent relative economic stagnation has no bearing on whether the forces making for convergence among industrial capitalist economies are strong. There is only one centrally planned economy in the unbiased sample, and its removal has negligible quantitative effects on the estimated degree of convergence.

[10] A strong case can be made for including Czechoslovakia and Hungary in this extended sample even though estimates of their 1870 per capita GNP fall just below that of Finland. Bohemia was industrializing at a pace equal to that of Austria (see Ivan Berend and György Ránki, 1974, David Good, 1984, 1986, and Nachum Gross, 1973). William Ashworth (1977) and John Komlos (1983) believe respectively that the areas that were to become Czechoslovakia and Hungary grew faster than the area around Vienna over 1870 and 1914. It is this rapid growth that, combined with Bairoch's (1981) estimate of the relative income gap between Austria on the one hand and Czechoslovakia and Hungary on the other, places their estimated 1870 per capita income below Finland's and so leads to their exclusion from the sample.

All the additional nations have strong claims to belong to the 1870 convergence club. All were well integrated into the Europe-based international economy. All had bright development prospects as of 1870. Saxony-Silesia was not much inferior to the Ruhr in industrialization.[11] Argentina, Chile, and New Zealand were grouped in the nineteenth century with Australia and Canada as countries with temperate climates, richly endowed with natural resources, attracting large-scale immigration and investment, and exporting large quantities of raw and processed agricultural commodities. They were all seen as natural candidates for the next wave of industrialization.[12]

Ireland's economy was closely integrated with the most industrialized economy in the world. Spain and Portugal had been the technological leaders of Europe during the initial centuries of overseas expansion—their per capita incomes were still above the European mean in the 1830s (Paul Bairoch, 1981)—and had retained close trading links with the heart of industrial Europe. Coke was used to smelt iron in Asturias in the 1850s, and by 1877 3,950 miles of railroad had been built in Spain. It is difficult to see how one could exclude Portugal and Spain from the convergence club without also excluding nations like Sweden and Finland.[13]

Baumol's sample failed to include those nations that should have belonged to any hypothetical convergence club but that nevertheless did not converge. The enlarged sample might include nations not in the 1870 convergence club. Consider Kuwait today: Kuwait is rich, yet few would take its failure to maintain its relative standard of living over the next fifty years as evidence against convergence. For Kuwait's present wealth does not necessarily carry with it the institutional capability to turn oil wealth into next generation's industrial wealth.

No nation in the once-rich twenty-two is in the same class as Kuwait.[14] The prosperity of the temperate settler colonies—Argentina, Chile, and New Zealand—was built on European capital, labor, and skills at least as much as on natural resources. And the economic links from East Germany, Ireland, Portugal, and Spain on the one hand to the world's industrial core on the other were all strong enough in 1870 to make them natural candidates for rapid industrialization.

The volume of overseas investment poured into the additional nations by investors from London and Paris between 1870 and 1913 tells us that investors thought these nation's development prospects good. Herbert Feis' (1930)[15] standard estimates of French and British overseas investment show the six non-European nations among the top ten[16]

---

[11] In fact, Saxony was more industrialized than the Ruhr until a surprisingly late data. By 1850 Saxony exported textiles, textile machinery, steam engines, and locomotives. Even in 1925 Saxony was still more densely settled than any country in the world. See Sidney Pollard (1981).

[12] The first essay of Carlos Díaz-Alejandro (1970) is very instructive. It attempts to recall the days before 1930 when Argentina would have indisputably been considered part of the first world. This point is also made by W. Arthur Lewis (1978).

[13] See Pollard (1981). Pollard makes the parallel between the situations of Iberia and Scandinavia explicit.

[14] Except possibly for Australia. Maddison's estimate of Australian 1870 per capita income places Australia 43 percent higher than the next most prosperous nation, Britain. Yet Australia did not have the most technologically sophisticated economy in the world in 1870. Australia's prosperity was built on its abundant and fertile land and the European demand for sheep, just as Kuwait's prosperity today is built on the European demand for oil. Should the failure of Australia to maintain its 50 percent per capita income edge over the cluster of next most prosperous nations count as strong evidence for convergence? I would suspect not, yet in the regressions Australia delivers the strongest single impact in favor of convergence of any nation. In the text, Australia is kept in the once-rich twenty-two sample in order to avoid the suspicion of *ex post* sample reselection. The Appendix considers the effect of removing Australia from the sample.

[15] Which still remains the standard source on overseas investment. See Michael Edelstein (1982). The only major point possibly at issue is whether the London capital market channeled just British or a much wider pool of savings into international investments.

[16] The foreign investment figures do provide a powerful argument for adding other Latin American nations —Mexico, Brazil, and Cuba—to the sample of those that ought to have been in the convergence club. Inclusion of these nations would weigh heavily against convergence.

TABLE 2—TOP TEN RECIPIENTS OF BRITISH AND FRENCH[a] OVERSEAS INVESTMENT PER CAPITA AS OF 1913

Nation	Pounds Invested Per Capita	Belongs to Once-Rich Twenty-Two Sample?
Canada	86	Yes
Australia	57	Yes
New Zealand[b]	57	Yes
Argentina	54	Yes
South Africa	27	No
Chile	26	Yes
Cuba	17	No
Mexico	10	No
Brazil	8	No
United States	8	Yes

[a] French investments in Latin America allocated to nations in the same proportions as British investments.
[b] Not distinguished from Australia.
Note: Investment estimates from Feis (1930). Population estimates from Mitchell (1975).

recipients of investment per capita from France and Britain, and four of the five top recipients of investment belong to the once-rich twenty-two.[17] Every pound or franc invested is an explicit bet that the recipient country's rate of profit will remain high and an implicit bet that its rate of economic growth will be rapid. The coincidence of the nations added on a per capita income basis and the nations that would have been added on a foreign investment basis is powerful evidence that these nations do belong in the potential convergence club.[18]

Errors in estimating 1870 income are unavoidable and produce equal and opposite errors in 1870–1979 growth. These errors therefore create the appearance of convergence where it does not exist in reality.[19] If 1870 income is measured with error, the appropriate regression model is not

$$(1)\quad (1979\text{ Income}) - (1870\text{ Income}) = \alpha + \beta(1870\text{ Income}) + \varepsilon_i,$$

where ε_i is a random error, but

$$(2)\quad (1979\text{ Income}) - (\text{True }1870\text{ Income}) = \alpha + \beta(\text{True }1870\text{ Income}) + \varepsilon_i$$

$$(3)\quad (\text{Estimated }1870\text{ Income}) = (\text{True }1870\text{ Income}) + \eta_i,$$

where η_i is a second random error and where true 1870 per capita income is not observed. The system composed of (2) and (3) is not identified under standard assumptions unless instruments are available. In this case instruments cannot be found: a variable correlated with 1870 income is one of the sources used to construct the estimates of 1870 income. The model is identified if one assumes that the errors ε and η are uncorrelated and fixes a value for the ratio of the error variances ρ (Edmond Malinvaud, 1966)[20]

$$(4)\qquad \rho = \left(\frac{\sigma_\eta^2}{\sigma_\varepsilon^2}\right).$$

---

[17] Japan would not merit inclusion in the 1870 convergence club on the basis of foreign investment before World War I, for Japanese industrialization was not financed by British capital. Foreign investors' taste for Japan was much less, investment being equal to about one pound sterling per head and far below investment in such nations as Venezuela, Russia, Turkey, and Egypt. Admittedly, Japan was far away and not well known. But who would have predicted that Japan would have five times the measured per capita GNP of Argentina by 1979?

[18] European outmigration was also overwhelmingly directed to nations in the once-rich twenty-two. Between 1860 and 1920, some 25 million (gross) departed Europe for the United States, some 5 million each for Canada and Argentina, more than 4 million for Australia and New Zealand, and more than 3 million for Brazil. See Ashworth (1987).

[19] By contrast, errors in measuring 1979 per capita income induce no systematic bias in the relationship between standard of living in 1870 and growth since, although they do diminish the precision of coefficient estimates.

[20] Letting y_i represent the log of estimated initial 1870 income, x_i represent the log of true initial 1870 income, and g_i represent estimated growth, the maximum likelihood estimates of "true" 1870 per capita income $\hat{x}_i$ and of the slope coefficient $\hat{\beta}$ solve the system

$$(5)\qquad (\hat{x})_i = \frac{(1+\rho+\rho\hat{\beta})y_i + \rho(1+\hat{\beta})g_i}{1+\rho(1+\hat{\beta})^2},$$

$$(6)\qquad \hat{\beta} = \frac{\Sigma(\hat{x})_i(y_i+g_i)}{\Sigma(\hat{x}_i)^2} - 1.$$

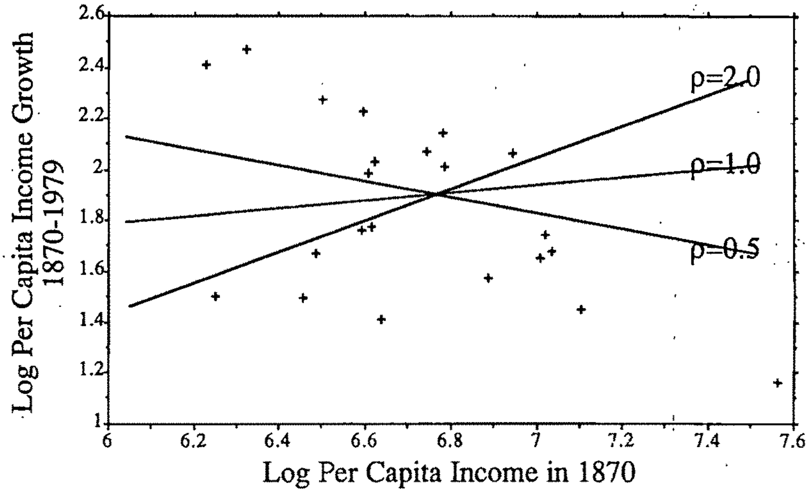


FIGURE 3. MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO, 1870–1979

TABLE 3—MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Standard Error of Regression	Standard Error in 1870 PCI
0.0	−.566	.144	.207	.0
0.5	−.292	.192	.192	.136
1.0	0.110	.283	.170	.170
2.0	0.669	.463	.134	.190
∞	1.381	.760	.0	.196

Note: Data from Maddison (1982) and as described in the Appendix.

A reasonable central value to fix for ρ is one, which implies that measurement error in 1870 income is equal in size to the regression disturbance. This if anything overestimates the precision of 1870 income estimates. Any believer in convergence is, moreover, committed to the position that ρ is large, for convergence requires not only a negative slope β but also a small variance of the regression disturbance ε. Nations have not converged if the spread of their relative incomes remains large, even if leaders now were followers a century ago.

Taking ρ equal to one as a central case is far more reasonable than the implicit fixing of ρ equal to zero of ordinary least squares. In Table 3 and Figure 3 results are reported for Baumol's regression, using the once-rich twenty-two sample over the period 1870–1979, for ρ equal to zero,[21] to one-half, to one (the central case), to two, and to infinity.[22] Reporting results for this range of ρ values summarizes how beliefs about data quality map into conclusions about convergence.

From one point of view, the relatively poor quality of much of the nineteenth century data is not a severe liability for this paper. Only if there is less measurement error than allowed for will the results be biased against convergence. A more direct check on the importance of measurement error can be performed by examining convergence starting at some later data for which income estimates are based on a firmer

---

[21] The ordinary least squares regression case in which measurement error in the independent variable is assumed nonexistent.

[22] This corresponds to the inverse regression in which ordinary least squares is used but the roles of dependent and independent variables are switched.

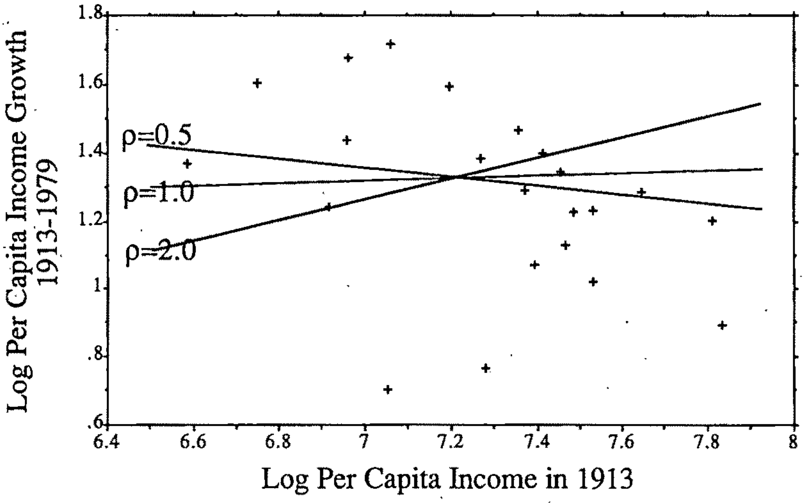


FIGURE 4. MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO, 1913–1979

foundation. A natural such data is 1913.[23] The relationship between initial income and subsequent growth is examined for the period 1913–1979 in Table 4 and Figure 4.

The longer 1870–1979 sample of Table 3 and Figure 3 is slightly more hospitable to convergence than is the 1913–1979 sample, but for neither sample do the regression lines reveal a significant inverse relationship between initial income and subsequent growth. When it is assumed that there is no measurement error in 1870 income, there is a large negative slope to the regression line. But even in this case the residual disturbance term is large. When measurement error variance is assumed equal to half disturbance variance, the slope is slightly but not significantly negative.

TABLE 4—MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO, 1913–1979

ρ	Slope Coefficient β	Standard Error of Slope	Standard Error of Regression	Standard Error in 1870 PCI
0.0	−.333	.116	.171	.0
0.5	−.140	.136	.151	.107
1.0	0.021	.158	.133	.133
2.0	0.206	.191	.106	.150
∞	0.444	.238	.0	.167

Note: Data from Maddison (1982) and as described in the Appendix.

For the central case of equal variances, growth since 1870 is unrelated to income in 1870. There is no convergence. Those countries with income edges, have on average maintained them. If measurement error is assumed larger than the regression disturbance there is not convergence but divergence. Nations rich in 1870 or 1913 have subsequently widened relative income gaps. The evidence can be presented in other ways. The standard deviations of log income are

[23] The data for 1913 are much more plentiful and solid than for other years in the early years of the twentieth century because of the concentration of historians' efforts on obtaining a pre-World War I benchmark. Beginning the sample at 1913 does mean that changes in country's "social capability" for development as a result of World War I appear in the error term in the regression. If those nations that suffered most badly in World War I were nations relatively poor in World War I, there would be cause for alarm that the choice of 1913 had biased the sample against finding convergence when it was really present. But the major battlefields of World War I lay in and the largest proportional casualties were suffered by relatively rich nations at the core of industrial Europe.

TABLE 5—STANDARD DEVIATIONS OF LOG OUTPUT FOR MADDISON'S SIXTEEN AND THE ONCE-RICH TWENTY-TWO

Sample	1870	1913	1979
Maddison's 16	.411	.355	.145
Once-Rich 22	.315	.324	.329

Note: Data from Maddison (1982) and as described in the Appendix.

given in Table 5. Maddison's sixteen do converge: the standard deviation of log income in 1979 is only 35 percent of its 1870 value. But the appearance of convergence is due to selection bias: the once-rich twenty-two have as wide a spread of relative incomes today as in 1870.

The failure of convergence to emerge for nations rich in 1870 is due to the nations —Chile, Argentina, Spain, and Portugal—in the bottom left corner of Figures 2 through 4. In the early 1970s none of these was a democracy. Perhaps only industrial nations with democratic political systems converge. A dummy variable for democracy over 1950–80 is significant in the central ($\rho=1$) case in the once-rich twenty-two regression in a at the 1 percent level, as detailed in Table 6.

But whether a nation is a democracy over 1950–80 is not exogenous but is partly determined by growth over the preceding century. As of 1870 it was not at all clear which nations would become stable democracies. Of the once-rich twenty-two, France, Austria (including Czechoslovakia), and Germany were empires; Britain had a restricted franchise; Spain and Portugal were semiconstitutional monarchies; the United States had just undergone a civil war; and Ireland was under foreign occupation. That all of these countries would be stable democracies by 1950 seems *ex ante* unlikely. Table 7 shows that shifting to an *ex ante* measure of democracy[24] removes the correlation. Whether a nation's politics are democratic in 1870 has little to do with growth since. The elective affinity of democracy and opulence is not one way with democracy as cause and opulence as effect.

There is one striking *ex ante* association between growth over 1870–1979 and a predetermined variable: a nation's dominant religious establishment. As Table 8 shows, a religious establishment variable that is one for Protestant, one-half for mixed, and zero for Catholic nations is significantly correlated with growth as long as measurement error variance is not too high.[25]

This regression is very difficult to interpret.[26] It does serve as an example of how

---

[24] Defined as inclusion in the electorate of more than half the adult male population.

[25] The once-rich twenty-two are split into nations that had Protestant religious establishments in 1870 (Australia, Denmark, Finland, E. Germany, Netherlands, New Zealand, Norway, Sweden, U.K., and United States), intermediate nations—nations that either were split in established religion in 1870 or that had undergone violent and prolonged religious wars between Protestants and Catholics in the centuries after the Protestant Reformation—(Belgium, Canada, France, West Germany, and Switzerland), and nations that had solid Catholic religious establishments in 1870 (Argentina, Austria, Chile, Ireland, Italy, Portugal, and Spain). This classification is judgmental and a matter of taste: are the Netherlands one of the heartlands of the Protestant Ethic or are they one of the few nations tolerant and pluralistic on matters of religion in the seventeenth century?

[26] The easy explanation would begin with the medieval maxim *homo mercator vix aut numquam placere potest Deo*: the merchant's business can never please God. Medieval religious discipline was hostile to market capitalism, the Protestant Reformation broke this discipline down in some places, and capitalism flourished most and modern economic growth took hold strongest where this breakdown of medieval discipline had been most complete.

But this easy explanation is at best incomplete. Initially the Reformation did not see a relaxation of religious control. Strong Protestantism—Calvin's Geneva or Cromwell's Republic of the Saints—saw theology and economy closely linked in a manner not unlike the Ayatollah's Iran. And religious fanaticism is not often thought of as a source of economic growth.

Nevertheless the disapproval of self-interested profit seeking by radical Protestantism went hand-in-hand with seventeenth century economic development. And by 1800 profit seeking and accumulation for accumulation's sake had become morally praiseworthy activities in many nations with Protestant religious establishments. How was the original Protestant disapproval for the market transformed? Accounting for the evolution of the economic ethic of the Protestant West from Jean Calvin to Cotton Mather to Benjamin Franklin to

TABLE 6—DEMOCRACY OVER 1950–1980 AND LONG-RUN GROWTH FOR THE ONCE-RICH TWENTY-TWO, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Coefficient on Democracy Variable	Standard Error	Standard Error in 1870 PCI	Standard Error of Regression
0.0	−.817	.277	.495	.085	.155	.0
0.5	−.744	.203	.476	.084	.154	.109
1.0	−.599	.208	.437	.090	.150	.150
2.0	0.104	.227	.248	.071	.131	.185
∞	1.137	.019	.044	.003	.0	.198

Note: Data from Maddison (1982) and as described in the Appendix.

TABLE 7—DEMOCRACY IN 1870 AND LONG-RUN GROWTH FOR THE ONCE-RICH TWENTY-TWO, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Coefficient on Democracy Variable	Standard Error	Standard Error in 1870 PCI	Standard Error of Regression
0.0	−.567	.342	.001	.091	.207	.0
0.5	−.272	.322	−.038	.094	.192	.136
1.0	0.164	.454	−.095	.115	.169	.169
2.0	0.742	.976	−.170	.180	.131	.155
∞	1.231	.167	−.195	.022	.0	.194

Note: Data from Maddison (1982) and as described in the Appendix.

TABLE 8—DOMINANT RELIGION IN 1870 AND LONG-RUN GROWTH FOR THE ONCE-RICH TWENTY-TWO, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Coefficient on Religion Variable	Standard Error	Standard Error in 1870 PCI	Standard Error of Regression
0.0	−.789	.252	.429	.088	.166	.0
0.5	−.688	.225	.403	.088	.164	.116
1.0	−.470	.248	.347	.098	.158	.158
2.0	0.375	.232	.132	.061	.132	.187
∞	1.199	.021	−.003	.004	.0	.197

Note: Data from Maddison (1982) and as described in the Appendix.

culture may be associated with substantial divergence in growth performance. But "Protestantism" is correlated with many things—early specialization in manufacturing (for a given level of income), a high investment ratio, and a northern latitude, to name three. Almost any view—except a belief in convergence—of what determines long-run growth is consistent with this correlation between growth and religious establishment. Moreover, this correlation will not last: neither fast grower Japan nor fast grower Italy owes anything to the Protestant ethic.[27] The main message of Table 8 is that,

---

Andrew Carnegie is a deep puzzle in economic history. The best analysis may still be the psychological account given by Max Weber (1958). Originally published in 1905.

[27] But see Michio Morishima (1982).

for the once-rich twenty-two, a country's religious establishment has been a surprisingly good proxy for the social capability to assimilate modern technology.

The long-run data do not show convergence on any but the most optimistic reading. They do not support the claim that those nations that should have been able to rapidly assimilate industrial technology have all converged. Nations rich among the once-rich twenty-two in 1870 have not grown more slowly than the average of the sample. And of the nations outside this sample, only Japan has joined the industrial leaders.

This is not to say that there are no forces pushing for convergence. Convergence does sometimes happen. Technology is a public good. Western Europe (except Iberia) and the British settlement colonies of Australia, Canada, and the United States are now all developed. Even Italy, which seemed outside the sphere of advanced capitalism two generations ago, is near the present income frontier reached by the richest nations. The convergence of Japan and Western Europe toward U.S. standards of productivity in the years after World World II is an amazing achievement, and this does suggest that those present at the creation of the post-World World II international order did a very good job. But others—Spain, Portugal, Ireland, Argentina, and Chile—that one would in 1870 have thought capable of equally sharing this prosperity have not done so.[28] The capability to assimilate industrial technology appears to be surprisingly hard to acquire, and it may be distressingly easy to lose.[29]

The forces making for "convergence" even among industrial nations appear little stronger than the forces making for "divergence." The absence of convergence pushes us away from a belief that in the long-run technology transfer both is inevitable and is the key factor in economic growth. It pushes us away from the belief that even the nations of the now industrial West will have roughly equal standards of living in 2090 or 2190. And the absence of convergence even among nations relatively rich in 1870 forces us to take seriously arguments like Romer's (1986) that the relative income gap between rich and poor may tend to widen.

APPENDIX

Estimates of 1979 per capita income for all nations in the sample are taken from Robert Summers and Alan Heston (1984), are calculated in Irving Kravis, Heston, and Summers' (1978) 1975 "international dollar" price measure. These 1979 estimates are not exactly equal to the estimates used by Maddison (1982) and Baumol (1986) for three reasons. First, Maddison (1982) relied on an earlier application of the United Nations International Comparison Project methodology (see Maddison, 1982; Kravis, Heston, and Summers, 1978; and Summers and Heston, 1984). Second, Maddison prefers to work with data in U.S. relative prices because the price structures of the other nations in his sample are becoming more and more like that of the United States as time passes. Third, Maddison (1982) works in 1970 prices but the natural index date for international dollars is 1975.

Since the price structures of some of the nations in the extended sample are still far from that of the United States, I find considerable merit in the Kravis et al. international dollar measure. In addition, the use of U.S. relative prices would significantly increase the relative spread of incomes in the sample; use of the international dollar is therefore the choice of price weights most favorable to "convergence."

Per capita income estimates for 1870 and 1913 for the nations of Maddison's sixteen are calculated using Summers and Heston's (1984) estimates of 1979 per capita income and Maddison's (1982) estimates of 1870–1979 and 1913–79 per capita income growth.[30]

---

[28] One can find good reasons—ranging from the Red Army to landlord political dominance to the legacy of imperialism—for the failure of each of the additional nations to have reached the world's achieved per capita income frontier in 1979. But the fact that there are good reasons for the relative economic failure of each of these seven nations cases substantial doubt on the claim that the future will see convergence, for "good reasons" for economic failure will always be widespread. It is a safe bet that in 2090 one will be able *ex post* to identify similar "good reasons" lying behind the relative economic decline of those nations that will have fallen out of the industrial core.

[29] Britain, the first industrial nation, is now almost as far behind today's industrial leaders in relative per capita income terms as nations like France and Germany were behind Britain in 1870. Something more than the logic of convergence may be at work.

[30] For an overview of the history of world development during the past century, see Pollard (1981) for Europe, Ashworth (1977), Arthur Lewis (1978), and W. W. Rostow (1978) for the world, and of course, Simon Kuznets (1966).

TABLE A1—MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO WITH CHILE SUBTRACTED, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Standard Error of Regression	Standard Error in 1870 PCI
0.0	−.723	.137	.186	.0
0.5	−.580	.175	.181	.128
1.0	−.231	.264	.169	.169
2.0	0.614	.604	.135	.191
∞	1.738	1.229	.0	.200

Note: Data from Maddison (1982) and as described in the Appendix.

This preserves Maddison's estimates of growth rates. It is obvious that error in estimating growth rates will induce error in estimated initial per capita income. Thus allowance is made, as discussed in the text, in the statistical work for errors in the independent variable.

Per capita income estimates for 1870 and 1913 for Ireland are calculated using the British per capita income estimates found in Maddison (1982) and the estimate of the relation between Irish and British per capita incomes found in R. C. O. Matthews, C. H. Feinstein, and J. C. Odling-Smee (1982; see also Feinstein, 1972). Matthews et al. estimate that Irish per capita income was 54 percent of British in 1913 and that Irish and British per capita incomes grew at indistinguishable rates over the 1870 to 1913 period.

Per capita income estimates for 1870 and 1913 for Argentina are taken from Maddison (1970). The 1870 estimate depends heavily on Carlos Díaz-Alejandro's (1970) judgment and should not under any circumstances be cited for any purposes dealing with Argentinian development alone. The estimate is sufficiently shaky to be unacceptable for such purposes, although it is barely acceptable as an estimate for a comparative project like this one in which omission of nations for lack of acceptable data is not an option—due to the potential generation of selection bias—and in which errors in variables are adequately handled from a statistical point of view. The 1913 estimate is based on a considerably more solid foundation.

Per capita income estimates for Chile in 1913 are taken from Maddison (1970). Estimates of Chilean per capita income in 1870 are considerably more difficult to construct. Markos Mamalakis (1976) cites estimates of agricultural production that suggest that the agricultural sector alone produced enough to give all Chileans an annual income of more than 220 1975 dollars in 1841. On the assumptions that per capita income growth was constant from 1841 to 1913 and that the agricultural sector was initially three quarters of the economy, Chilean per capita income growth before 1913 was some 1.3 percent per annum, and 1870 per capita income was around 519 1975 dollars. This estimate is perhaps the shakiest of all, and places Chile close to the cutoff for inclusion in the sample. Although the volume of investment from London and Paris in Chile before 1913 strongly militates against Chile's exclusion, Table A1 reports quantitative statistical results for the basic regressions with Chile removed from the once-rich twenty-two sample.

Per capita income estimates for 1870 and 1913 for East Germany are calculated from data given in Maddison (1982), who reports German per capita income both not adjusting for changes in boundaries and adjusting for changes in boundaries (that is, calculating per capita income within the borders of present-day West Germany). "East Germany" before World War II therefore does not have the same boundaries as post-World War II East Germany. The bias introduced by including parts of what is now Poland in the area of East Germany before World War II should be small, for the additional areas include both industrial Silesia and agricultural Prussia.

Per capita income estimates for 1870 and 1913 for New Zealand,[31] Portugal, and Spain are derived from Bairoch (1976, 1981). Estimates of nineteenth century per capita income given by Bairoch are for 1860 and not, as in Maddison, for 1870. Bairoch's estimates are not in complete agreement with those of Maddison (1982), as is shown in Figure A1. The correlation between the two sets of mid-nineteenth century estimates of the log of per capita income for Maddison's sixteen is .91, and the estimated slope coefficients for the direct and reverse regressions of Maddison's estimates on Bairoch's are 1.16 and .71, respectively.

Maddison's estimates show a higher variance of log 1870 per capita income and are thus more favorable to the convergence hypothesis. For this reason, because Maddison's documentation is more complete, and because I am more comfortable with his estimates, past per capita income estimates are retained on Maddison's basis. The 1870 per capita income estimates for New Zealand, Portugal, and Spain are obtained by using the least squares regression of Maddison's estimates on

---

[31] See J. A. Dowie (1966). Whether New Zealand should be counted as a nation is open to debate: only 300 thousand people lived in New Zealand in 1870. Since New Zealand had an 1870 level of per capita income above the sample mean and has had relatively slow growth since, its inclusion in the sample is not hostile to convergence.

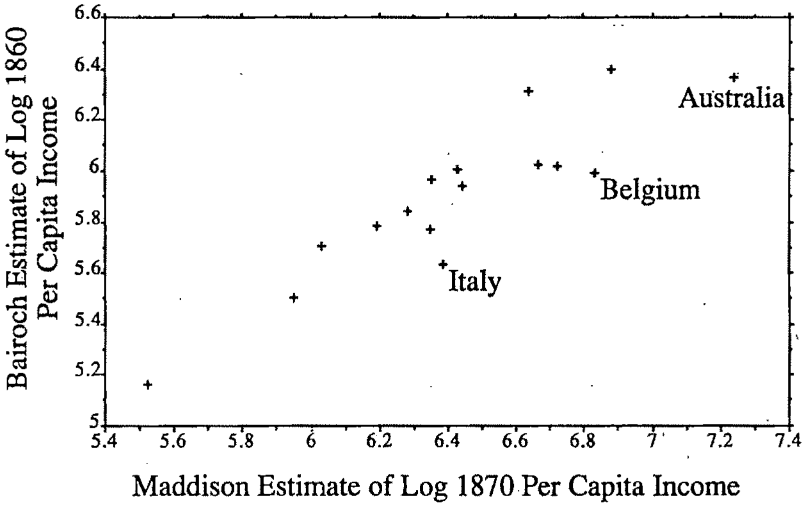


FIGURE A1. MADDISON AND BAIROCH ESTIMATES OF NINETEENTH CENTURY PER CAPITA INCOME

TABLE A2—MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO WITH AUSTRALIA OMITTED, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Standard Error of Regression	Standard Error in 1870 PCI
0.0	−.430	.182	.208	.0
0.5	0.157	.290	.181	.128
1.0	0.769	.459	.147	.147
2.0	1.297	.650	.110	.155
∞	1.677	.779	.0	.161

Note: Data from Maddison (1982) and as described in the Appendix.

Bairoch's to "predict" what Maddison would have estimated 1870 and 1913 per capita income to be had these nations fallen into his sample.

For the most part, minor adjustments of the sample by including or excluding individual nations make no perceptible difference in the quantitative results. Australia and Japan are the only exceptions, for each of these Asian outliers alone has the potential to cause noticeable shifts in the estimated degree of convergence by its inclusion or exclusion.

Australia is an especially troublesome case because it appears very possible that Australian income in 1870 may have been significantly lower than reported by Maddison. Bairoch's disagreement with Maddison has already been noted. Ian McLean and Jonathan Pincus (1983) believe that Australian 1870 per capita GNP has traditionally been overstated by as much as 33 percent. David Pope (1984) reports that, on the standard estimates that serve as a basis for Maddison (1982), Australian real wages fell between 1870 and 1920. The 50-year stagnation of real incomes found in Maddison's estimates suggests that high Australian incomes in 1870 bore little relation to the dynamics of the industrial revolution and of technology-driven economic growth.

The sensitivity of the empirical analysis to the inclusion of Australia is illustrated by Table A2, which reports regression results for the once-rich twenty-two sample with Australia omitted. Omitting Australia leaves a negative slope only in the case where there is not measurement error in 1870 per capita income, and in the central case where $\rho = 1.0$ creates a large degree of divergence. To make the point another way, consider the standard deviations of the once-rich twenty-two sample including and excluding Australia. Including Australia, the ratio of the standard deviations of 1979 to 1870 income is equal to 1.06. Excluding Australia, to 1.31. What would be substantial divergence in the relative spread of incomes with Australia excluded becomes no change in the relative spread with Australia included.

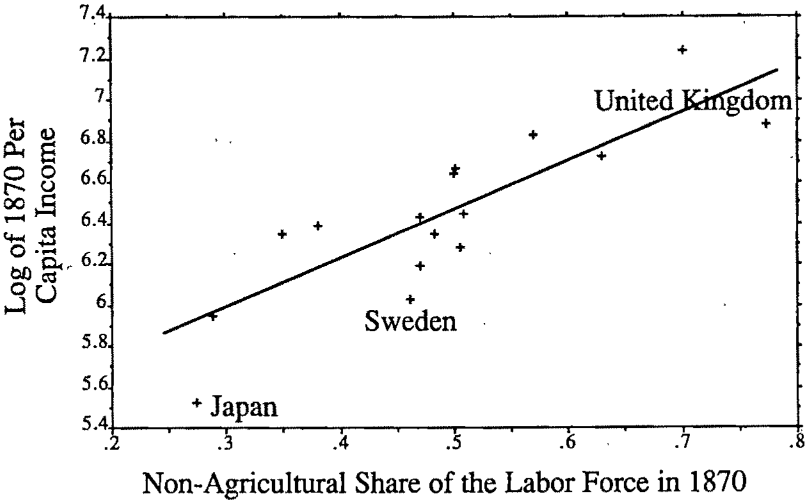


FIGURE A2. 1870 PER CAPITA INCOME AND NONAGRICULTURAL SHARE OF THE LABOR FORCE FOR MADDISON'S SIXTEEN

TABLE A3—MAXIMUM LIKELIHOOD ESTIMATION FOR THE ONCE-RICH TWENTY-TWO WITH JAPAN ADDED, 1870–1979

ρ	Slope Coefficient β	Standard Error of Slope	Standard Error of Regression	Standard Error in 1870 PCI
0.0	−.718	.124	.211	.0
0.5	−.580	.154	.205	.145
1.0	−.314	.216	.193	.193
2.0	0.373	.449	.158	.224
∞	1.444	.973	.0	.237

Note: Data from Maddison (1982) and as described in the Appendix.

The addition of Japan to the once-rich twenty-two has an effect roughly equal (in an opposite direction) to the subtraction of Australia, as Table A3 reveals. Although there is no *ex ante* warrant for including Japan in a sample of nations likely to rapidly industrialize as of 1870, its addition shifts the quantitative results toward showing convergence.

Looking back one can find structural features that make Japan's rapid industrial success less of a surprise. Late Tokugawa Japan was a substantially urban, commercialized society. It had a relatively small proportion of its labor force in agriculture given its reported per capita income. According to Maddison, Japan had a proportion of the labor force in agriculture not that different from Austria and Italy and almost exactly equal to Finland.[32] Yet according to Maddison Japanese per capita income in 1870 is only some 40 percent of per capita income in these first nations and less than two thirds per capita income in Finland. As Figure A2 demonstrates, the relatively low share of the labor force in agriculture given Japan's estimated per capita income casts doubt on the accuracy of the per capita income estimates. Japan does not fit the relation between nonagricultural labor force share and per capita income that holds for the rest of Maddison's sixteen.[33] Throughout most of the nineteenth century a larger

---

[32] And thus much below the agricultural share of nations with approximately Japan's 1870 per capita income.

[33] There are two other outliers that have a much lower proportion of their labor force in agriculture than one would think given their level of per capita income: Britain and Sweden. The reason for Britain's low share of the labor force in agriculture is obvious. The case of Sweden may be more akin to that of Japan; in Sweden also the social indicators suggest a much higher living standard than do the per capita income estimates.

TABLE A4—PER CAPITA INCOME ESTIMATES IN 1975 DOLLARS

Nation	Per Capita Income (1870)	Per Capita Income (1913)	Per Capita Income (1979)
Maddison's Sixteen			
Australia	1922	2523	6160
Austria	751	1436	5731
Belgium	1137	1778	6078
Canada	881	2085	7527
Denmark	883	1724	6621
Finland	506	1053	5640
France	847	1658	6705
W. Germany	731	1562	6789
Italy	746	1051	4424
Japan	328	621	5749
Netherlands	1104	1591	5778
Norway	665	1162	6475
Sweden	557	1336	6594
Switzerland	1118	1866	6388
UK	1214	1864	5166
United States	1038	2462	8205
Nation	**Per Capita Income (1870)**	**Per Capita Income (1913)**	**Per Capita Income (1979)**
Additional Seven Nations			
Argentina	762	1450	3119
Chile	519	1156	2337
E. Germany	741	1749	5409
Ireland	656	1007	3491
New Zealand	981	1624	4724
Portugal	637	725	2845
Spain	728	854	4246

Note: Data from Maddison (1982) and as described in the Appendix.

proportion of the Japanese population lived in the Shogun's seat of Edo than the proportion of the British population that lived in London (Edward Seidensticker, 1983). Japanese life expectancy was high. All of these social indicators fit badly with a level as low as the 300 1975 dollars a year used by Maddison. They suggest that per capita income may be significantly mismeasured, for social indicators suggest higher standards of living than do per capita income estimates.[34]

Table A4 presents the per capita income estimates used in this paper.

---

[34] See Susan Hanley (1983), Yasukichi Yasuba (1986), but also Kazushi Ohkawa and Henry Rosovsky (1973). If per capita income was not underestimated, then Japan's ability to assimilate industrial technology was very advanced for its wealth. See G. C. Allen (1981).

REFERENCES

Abramovitz, Moses, "Catching Up, Forging Ahead, and Falling Behind," *Journal of Economic History*, June 1986, *46*, 385–406.

Adelman, Irma and Morris, Cynthia Taft, "Patterns of Industrialization in the Nineteenth and Early Twentieth Centuries," in Paul Uselding, ed., *Research in Economic History*, 1980, Vol. 5, Greenwich: JAI Press, 217–46.

Allen, G. C., *The Japanese Economy*, London: Weidenfeld and Nicolson, 1981.

Ashworth, William, *A Short History of the International Economy Since 1850*, 4th ed., London: Longman Group, 1987.

______, "Typologies and Evidence: Has Nineteenth Century Europe a Guide to Economic Growth?," *Economic History Review*, February 1977, *30*, 140–58.

Bairoch, Paul, "The Main Trends in National Economic Disparities Since the Industrial Revolution," in Paul Bairoch and Maurice Lévy-Leboyer, eds., *Disparities in Economic Development Since the Industrial Revolution*, New York: St. Martin's Press, 1981.

______, "Europe's Gross National Product, 1800–1975," *Journal of European Economic History*, Summer 1976, *5*, 273–340.

Baumol, William, "Productivity Growth, Convergence, and Welfare," *American Economic Review*, December 1986, *76*, 1072–85.

______, "America's Productivity 'Crisis'," *The New York Times*, February 15, 1987, 3:2.

Berend, Ivan and Ránki, György, *Economic Development in East Central Europe in the Nineteenth and Twentieth Centuries*, New York: Columbia University, 1974.

Clark, Gregory, "Why Isn't the Whole World Developed? Lessons from the Cotton Mills," *Journal of Economic History*, March 1987, *47*, 141–74.

Díaz-Alejandro, Carlos, *Essays on the Economic History of the Argentine Republic*, New Haven: Yale, 1970.

Dowie, J. A., "The Course and Character of Capital Formation in New Zealand, 1871–1900," *New Zealand Economic Pa-*

pers, 1966, *1*, no. 1, 38–58.

Easterlin, Richard, "Why Isn't the Whole World Developed?," *Journal of Economic History*, March 1981, *41*, 1–19.

Edelstein, Michael, *Overseas Investment in the Age of High Imperialism*, New York: Columbia University, 1982.

Feinstein, Charles, *National Income, Output, and Expenditure in the United Kingdom*, Cambridge, U.K.: Cambridge University Press, 1972.

Feis, Herbert, *Europe, the World's Banker*, New Haven: Yale, 1930.

Gerschenkron, Alexander, *Economic Backwardness in Historical Perspective*, Cambridge: Harvard University, 1962.

Good, David, "Uneven Development in the Nineteenth Century," *Journal of Economic History*, March 1986, *46*, 137–51.

______, *The Economic Rise of the Habsburg Empire*, Berkeley: University of California, 1984.

Gross, Nachum, "The Industrial Revolution in the Habsburg Monarchy," in Carlo Cipolla, ed., *The Fontana Economic History of Europe*, London: Fontana, 1973.

Hanley, Susan, "A High Standard of Living in Nineteenth Century Japan: Fact or Fantasy?," *Journal of Economic History*, March 1983, *43*, 183–92.

Komlos, John, *The Habsburg Monarchy as a Customs Union*, Princeton: Princeton University, 1983.

Kravis, Irving, Heston, Alan and Summers, Robert, *The United Nations International Comparison Project: Phase II: International Comparisons of Real Product and Purchasing Power*, Baltimore: Johns Hopkins University, 1978.

Kuznets, Simon, *Modern Economic Growth: Rate, Structure, Spread*, New Haven: Yale, 1966.

Lewis, W. Arthur, *Growth and Fluctuations, 1870–1913*, London: Allen & Unwin, 1978.

Maddison, Angus, "Growth and Slowdown in Advanced Capitalist Economies," *Journal of Economic Literature*, June 1987, *25*, 649–98.

______, "A Comparison of Levels of GDP per Capita in Developed and Developing Countries, 1700–1980," *Journal of Economic History*, March 1983, *43*, 27–41.

______, *Phases of Capitalist Development*, Oxford: Oxford University Press, 1982.

______, *Economic Progress and Policy in Developing Countries*, London: Allen & Unwin, 1970.

______, *Economic Growth in the West*, New York: The Twentieth Century Fund, 1964.

Malinvaud, Edmond, *Statistical Methods of Econometrics*, Amsterdam: North-Holland, 1966.

Mamalakis, Markos, J., *The Growth and Structure of the Chilean Economy*, New Haven: Yale, 1976.

Marx, Karl, "The Future Results of British Rule in India," in *Surveys from Exile*, New York: Penguin, 1973. Originally published in 1853.

Matthews, R. C. O., Feinstein, C. H. and Odling-Smee, J. C., *British Economic Growth 1856–1973*, Stanford: Stanford University, 1982.

McLean, Ian and Pincus, Jonathan, "Did Australian Living Standards Stagnate between 1890 and 1940?," *Journal of Economic History*, March 1983, *43*, 193–202.

Mill, John Stuart, *Principles of Political Economy, with Some of Their Applications to Social Philosophy*, New York: Penguin, 1970. Originally published in 1848.

Mitchell, Brian, *European Historical Statistics*, London: Macmillan, 1975.

Morishima, Michio, *Why Has Japan, 'Succeeded'?*, Cambridge: Cambridge University Press, 1982.

Ohkawa, Kazushi and Rosovsky, Henry, *Japanese Economic Growth*, Stanford: Stanford University, 1973.

Pollard, Sidney, *Peaceful Conquest*, Oxford: Oxford University Press, 1981.

Pope, David, "Rostow's Kondratieff Cycle in Australia," *Journal of Economic History*, September 1984, *44*, 729–53.

Romer, Paul, "Increasing Returns and Long Run Growth," *Journal of Political Economy*, October 1986, *94*, 1002–37.

Rostow, W. W., *The World Economy: History and Prospect*, Austin: University of Texas, 1978.

Rudolph, Richard, *Banking and Industrializa-*

tion in Austria-Hungary, Cambridge: Cambridge University, 1976.

Seidensticker, Edward, *Low City, High City*, New York: Knopf, 1983.

Summers, Robert and Heston, Alan, "Improved International Comparisons of Real Product and Its Composition, 1950–1980," *Review of Income and Wealth*, Summer 1984, *30*, 207–62.

Weber, Max, *The Protestant Ethic and the Spirit of Capitalism*, New York: Scribner's, 1958. Originally published in 1905.

Yasuba, Yasukichi, "Standards of Living in Japan Before Industrialization," *Journal of Economic History*, March 1986, *46*, 217–24.

Productivity Growth, Convergence, and Welfare: Reply

By WILLIAM J. BAUMOL AND EDWARD N. WOLFF*

Preamble: *Mea Culpa*. Bradford De Long has succeeded in showing the error of my ways. I must agree with almost everything he says in his paper—except for his conclusion. This remark is meant neither to denigrate his results nor to exonerate my work, but to show (as I believe, from correspondence, DeLong now concurs) that his and our findings are compatible, and yield richer insights than were available before. An extensive study by Edward Wolff and myself indicates that there has indeed been convergence among productivity levels and per capita income levels of the richest industrialized countries, but the story is more complex than I had recognized.

First, let me agree that the sample of countries in my earlier article was inadvertently biased toward a showing of convergence. By using readily accessible data that dealt only with countries that afterward turned out to be successful I loaded the dice toward an appearance of convergence. It is no excuse (indeed perhaps the reverse) that I had recognized the problem in advance (under Paul Romer's prodding) and had said so in a footnote. DeLong has found estimates of early GNP per capita for a wider spectrum of countries (Paul Bairoch, 1976, and Angus Maddison, 1969) and uses them to show masterfully that a less biased sample exhibits little sign of convergence.[1]

How can this be consistent with a convergence story? Wolff and I set out to study *every* pertinent datum we could find and to subject the data to every relevant convergence test (virtuous data mining?). The results indicate that smaller groups of countries began to converge as early as, perhaps, 1860; that the size of the convergence club has since risen; but that for groups as large as DeLong's, for the period he studies, there was, as he concludes, probably no convergence. The following is a small representative sample of our many tests, with their (rather consistent) results. The full study will appear in our book, *Productivity Performance*: *The Long View*, forthcoming 1989 [W.J.B.].

Estimates for Earlier Periods and Their Implications. We used Bairoch's (1976) estimates of GNP per capita for 19 European countries, essentially by decade, from 1830 to 1913 to rank his countries in descending order of GNP per capita in 1870 (the starting point in our earlier study[2]), we constructed a sample of countries consisting of the set of eight countries at the top of the list. Then we successively constructed samples of the top 9, the top 10, and so on, until we got to the top 14 (beyond that number of countries too many data points were missing). For each sample size we calculated a time-series of the coefficients of variation (i.e., standard deviation divided by the mean)

---

*William J. Baumol is Professor of Economics, Princeton and New York Universities. Edward N. Wolff is Professor of Economics, New York University. The authors are very grateful for generous support of the research underlying this paper to the Division of Information Science and Technology of the National Science Foundation, to the Exxon Education Foundation, and the C. V. Starr Center for Applied Economics. We are also indebted to Vacharee Devakula for her very valuable assistance.

[1] At least one methodological problem does beset DeLong's calculations—its mixing of historical statistical estimates from different sources. As Summers has emphasized to us, and as our own calculations with more recent figures dramatically confirmed, two good time-series somewhat differently constructed can yield mutually consistent stories and yet produce a garbled tale when mixed together. However, neither this nor other possible criticisms on details undermines the basic validity of DeLong's central argument.

[2] We also used 1830, Bairoch's initial year. However, it is not obvious that an *ex ante* sample, with its bias in the other direction is what is wanted, either. Perhaps a proper test of convergence requires some intermediate starting date.

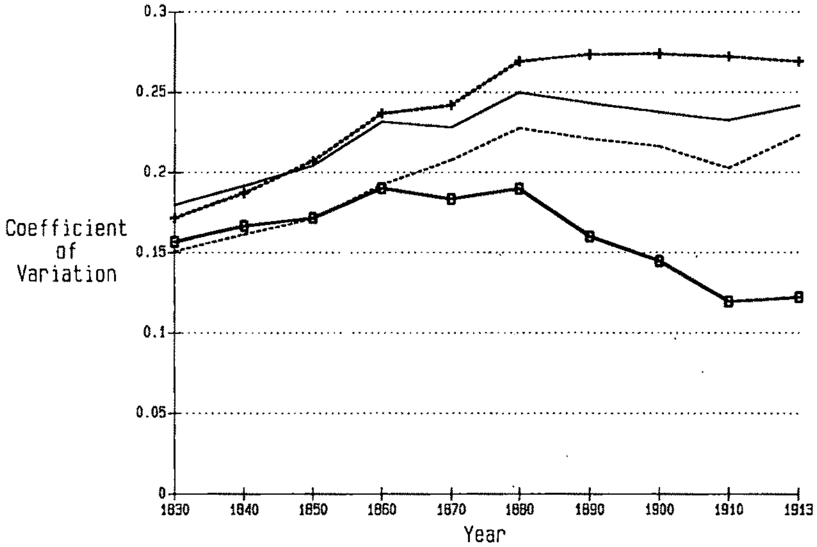


FIGURE 1. COEFFICIENT OF VARIATION GNP PER CAPITA, 1830–1913. BAIROCH COUNTRIES, 1870 RANKING. ⊖ TOP 8 COUNTRIES; --- TOP 9 COUNTRIES; — TOP 10 COUNTRIES; + TOP 11 COUNTRIES

for each year for which estimates were provided.

The results showed a straightforward pattern (Figure 1). Up to about 1860, for all samples, we found growing *divergence* in GDP per capita (rising coefficient of variation). But in the top group (8 countries) convergence appears to begin as early as 1860, and certainly by 1880. Slower and later convergence is shown for the top 9 and 10 country groups. However, if the remaining countries in Bairoch's table are added to the sample, divergence among this larger group continues all the way to the end of Bairoch's period.

These calculations, then, suggest that much of the nineteenth century was a period of divergence in standards of living of the leading European economies. Then, sometime toward the end of the century, this process began to erode, and was replaced by convergence among increasingly large sets of the initially (or later) more affluent of the countries. Moses Abramovitz has suggested to us that this is precisely what should have been expected. Before the Industrial Revolution the countries of Europe (with, perhaps, the exception of the Dutch Republic) were relatively homogeneous in their general poverty. Then Great Britain pulled ahead, inaugurating a growth in heterogeneity which was intensified as a small set of European leaders, including Belgium, Switzerland, the Netherlands, France, and Germany, also jumped ahead of the others. Only toward the end of the century was the leaders' example learned by others, so that convergence could extend to a group of any considerable size.

The Postwar Summers–Heston Data. Because Bairoch's figures stop in 1913, and his sample of countries is still far too small, they are insufficient to permit any conclusive *ex ante* classification test of the reality and prevalence of convergence. Fortunately, for 1950–80 Robert Summers and Alan Heston (1984) provide excellent data on real GDP (RGDP) per capita (in 1975 "international dollars") for 72 countries. We used these figures to test the hypothesis of our earlier study that among wealthier countries homogenization had occurred, but not among poorer countries. We tested this by calculating whether initially poorer countries subsequently grew faster than initially richer ones,

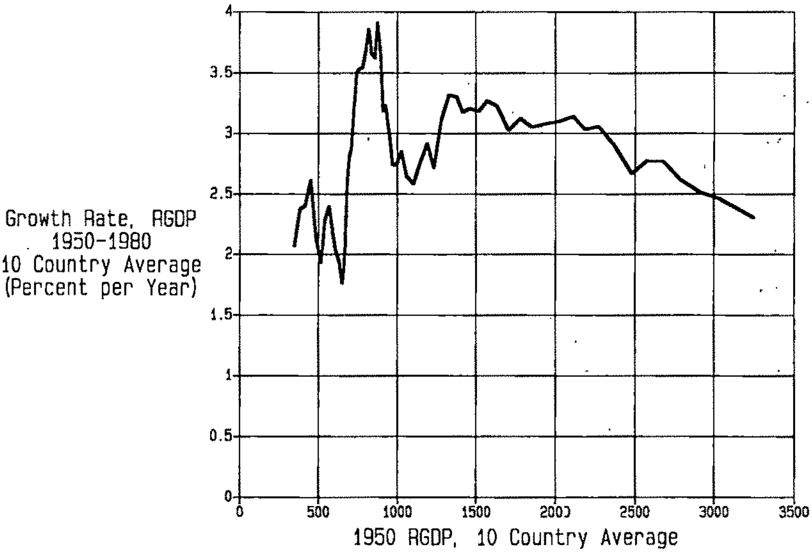


FIGURE 2. GROWTH RATE, RGDP (1950–80) VS. 1950 RGDP 10-COUNTRY MOVING AVERAGE, 72 COUNTRIES RANKED BY 1950 RGDP (SUMMERS–HESTON DATA)

as convergence requires. To reduce scatter and obtain a coherent pattern, we calculated moving averages for sets of 10 countries at a time. That is, the first set was made up of the 10 countries ranked lowest in terms of 1950 GDP per capita. The second set added the eleventh lowest and removed the (first) lowest, etc. For each group we graphed average 1950 RGDP on the horizontal axis, and the growth rate of RGDP, 1950 to 1980, on the vertical axis (Figure 2). The graph is an automatic *ex ante* stratification, since poorer countries in 1950 must lie to the left, while initially wealthier countries lie to the right. Our hypothesis amounts to the conjecture that the moving average graph would be roughly positively sloping toward the left (poorer countries growing more slowly), and distinctly negatively sloping toward the right, meaning that per capita incomes among LDCs, selected *ex ante*, had diverged, while the opposite had been true among initially "industrialized" countries.

This is, roughly, what Figure 2 shows. For nations with initial per-capita real GDPs below about $700, the curve's slope is highly erratic, and can, perhaps, be interpreted to be positive overall. Beginning possibly with a 700 dollar annual figure, and certainly above $1,300, the slope is clearly negative. Of the 72 countries, something between 29 and 52 fall in that group in the downward-sloping portion of the curve, while between 20 and 43 fall in the more or less positively sloping segment. This suggests that somewhere near the median in our sample of countries "the advantages of backwardness" do indeed begin to overbalance the counteracting forces, sociological, educational, and other.

In a more formal regression analysis, both a nonlinear relationship and a piecewise-linear relation composed of two line segments were fitted to the RGDP data.

The quadratic regression yielded the equation

$$\begin{aligned}\ln \text{RATIO} = {} & \underset{[4.2]}{0.586} + \underset{[2.1]}{0.00038}\,\text{RGDP50} \\ & - \underset{[2.2]}{(9.9/10^7)}\,\text{RGDP50}^2, \\ & \qquad R^2 = 0.07,\ N = 72,\end{aligned}$$

where RATIO = 1980 RGDP divided by

1950 RGDP and RGDP50 = per capita 1950 GDP. The maximum of the equation occurs at a 1950 RGDP value of about $1900. Both variables are significant at the 5 percent level, with the predicted signs. The results again show divergence among the 1950 lower-income countries and convergence among the higher-income ones.

The piecewise-linear regression was designed to attain its maximum near that of the nonlinear one, putting 17 of the 72 countries into the *ex ante* upper-income category. The resulting regression equation is

$$\begin{aligned} \ln \mathrm{RATIO} = {} & \underset{[5.8]}{0.658} + \underset{[1.9]}{0.00019}\ \mathrm{RGDP50} \\ & - \underset{[2.2]}{0.00044}\ \mathrm{D1900}, \\ & R^2 = 0.07,\ N = 72, \end{aligned}$$

where $\mathrm{D1900} = \mathrm{RGDP50}$ if $\mathrm{RGDP50} \geq \$1{,}900$ and 0 if $\mathrm{RGDP50} < \$1{,}900$. The first variable is significant at the 10 percent level, the second at the 5 percent level, and both have the predicted signs, confirming the quadratic-equation results. Finally, separate regressions were run for the upper-income sample and for the lower-income group.

For the upper-income group

$$\begin{aligned} \mathrm{RATIO} = {} & \underset{[7.7]}{3.3} - \underset{[12.5]}{0.00038}\ \mathrm{RGDP50}, \\ & R^2 = 0.30, \quad N = 17, \end{aligned}$$

and for the lower-income group

$$\begin{aligned} \mathrm{RATIO} = {} & \underset{[5.5]}{2.1} + \underset{[1.3]}{0.0005}\ \mathrm{RGDP50}, \\ & R^2 = 0.03, \quad n = 55. \end{aligned}$$

This provides strong evidence of convergence in the upper-income group but much weaker evidence of divergence among the lower-income countries.

The preceding tests can be criticized because they compare only 1950 and 1980, with no attention to intermediate year figures. It can be argued that RGDPs in 1950 were atypically diverse because of the great war damage to several industrialized economies. Their recovery naturally contributed abnormal convergence and thereby biased the calculation in the direction of our hypothesis.

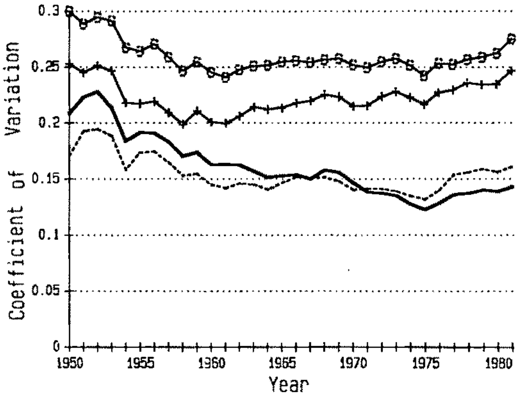


Figure 3. Coefficient of Variation, RGDP, 1950–1981, for Sets of Countries by 1950 RGDP Rank. --- Top 10 Countries; — Top 14 Countries; + Top 20 Countries; ⊖ Top 24 Countries

To avoid such problems, we calculated for different *ex ante* country samples the time path of the coefficient of variation *for each year* in the period 1950–1981. The Summers–Heston countries were again ranked on a 1950 basis. The time-series of coefficient of variation was calculated for the top 10 countries and for the top 12, 14, 16, 18, 22, 24, 26, 28, 30, 35, 40, 45, 50, 55, and 60 countries.

There is a sharp break in pattern of behavior between the samples that include fewer than 16 countries and those that include 16 or more. In Figure 3, the curve for the sample of the top 14 countries is typical for the (smaller) sets, containing the countries with the highest RDGP values in 1950. The coefficient of variation fell steadily and sharply throughout the period, except at its very beginning and very end. Noteworthy is the fairly steady but fairly modest rise since 1975. For larger samples divergence (positive slope) begins much earlier and continues far longer. In Figure 3, the curve for the sample of the top 26 countries, with its steady rise in divergence since 1961, is not atypical (though for larger samples patterns become more erratic). The coefficient of variation in 1981 was still about midway between its 1950 high and its 1961 low, so that about half of the initial homogenization has been lost.

These, then, are the facts, so far as they can be ascertained from the available data and many tests that lack of space prevents our reporting; indeed, from every test we could think of. They indicate that for perhaps the top 15 countries convergence has been marked and unambiguous, with a mild recent retreat. All countries together, excluding LDCs, have also shown some convergence. Yet, larger samples do not display convergence, in part because of the heterogeneous performance of the LDCs and failure of South American countries to live up to their growth promise.

This record offers no certain portents for the future, and it leaves open the most difficult and most crucial issue—just how countries achieve membership in the convergence club, and on what basis they are sometimes ejected.

REFERENCES

Abramovitz, Moses, "Catching Up, Forging Ahead, and Falling Behind," *Journal of Economic History*, June 1986, *46*, 385–406.

Bairoch, Paul, "Europe's Gross National Product, 1800–1973," *Journal of European Economic History*, Fall 1976, *5*, 213–340.

Baumol, William J., "Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show," *American Economic Review*, December 1986, *76*, 1072–85.

______, **Blackman, Batey, Sue-Anne and Wolff, Edward N.,** *Productivity Growth: the Long View*, Cambridge: MIT Press, forthcoming 1989.

De Long, Bradford, "Productivity Growth, Convergence, and Welfare: Comment," *American Economic Review*, December 1988, *78*, 1138–54.

Maddison, Angus, *Phases of Capitalist Development*, New York: Oxford University Press, 1982.

______, *Economic Growth in Japan and the USSR*, London: Allen & Unwin, 1969.

Romer, Paul M., "Increasing Returns and Long Run Growth," *Journal of Political Economy*, October 1986, *94*, 1002–37.

Summers, Robert and Heston, Alan, "Improved International Comparisons of Real Product and its Composition, 1950–1980," *Review of Income and Wealth*, June 1984, *30*, 207–62.

Wolff, Edward N., "Capital Formation and Long-Term Productivity Growth: A Comparison of Seven Countries," paper presented at the *Conference on Demographic Change and Economic Development*, Hagan, Germany, September 1986.

Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*: Comment

By JOHN J. MERRICK, JR.*

I have been able to replicate nearly perfectly the results of my May 1983 *Journal of Money, Credit and Banking* paper, "Financial Market Efficiency, the Decomposition of 'Anticipated' versus 'Unanticipated' Money Growth, and Further Tests of the Relation Between Money and Real Output."[1] I also have been able to reproduce the results obtained by William Dewald, Jerry Thursby, and Richard Anderson (hereafter, DTA) in their unsuccessful attempt to replicate my work as reported in their paper, "Replication in Empirical Economics: The *Journal of Money, Credit and Banking Project*" (September 1986 issue of this *Review*). Table A1 of the Appendix provides my evidence.[2] My analysis leads me to the conclusion that the data described and results originally reported in my 1983 paper are correct. The results obtained by DTA were caused by a data misalignment problem induced by my own error in labeling my stock return data in my post-acceptance data submission to the *JMCB* project.[3]

In their footnote 2, DTA state that "...we endeavored to keep (authors) informed of our replication efforts." For the record, I had never seen the results presented by DTA prior to their publication. DTA have since informed me that a draft of their paper was mailed to me in early 1985. Unfortunately, that draft never arrived and so the necessary detective work to uncover my reporting error was done after the fact.

Ironically, my reporting error would have been discovered back in 1984 had I responded completely to a data check request issued by another replicator, whose query was channeled to me by a member of DTA. This second replicator, who predated DTA and worked independently of them, generated results which differed dramatically from those of my 1983 paper and he specifically questioned my hand-labeled stock return data as the possible cause. However, his exact estimates (for example, insignificant autoregressive coefficients for quarterly money growth) led me to suggest to him (correctly) that important errors in his transcription of my money growth data were causing his most significant problems. Nevertheless, in suggesting the mislabeling possibility (a lead that I did not find time to follow up), this second replicator showed good intuition for what could go wrong in data exchanges supporting replication attempts.[4]

---

*Associate Professor of Finance, New York University Graduate School of Business Administration, 100 Trinity Place, New York, NY 10006.

[1] The reason why my replication attempt did not yield precisely the same results as my 1983 regression is not clear. Possibly, the estimation routines in the TSP package I used in the replication attempt differ slightly from those of the IDA package I employed for that 1983 paper. Also, the analyses were performed on different computer systems.

[2] DTA slightly misrepresented my 1983 paper's estimates. I reported coefficient standard errors, while DTA present computed t-ratios in their Table 6 (apparently using my estimated slope coefficients and standard errors as inputs). Such t-ratios are not directly comparable with (what I assume to be) DTA's own directly estimated t-ratios because of differences in rounding errors. Estimated standard errors should have been reported in the replication attempt (see Table A1).

[3] The section of my paper "sampled" by DTA concerned a money growth anticipations equation estimated for quarterly data for the 1947:3–1978:1 time period. The data misalignment problem which would have appeared in the project replicator's data file (once transcribed verbatim from my mislabeled printouts) concerned the lagged return on the Standard & Poor's 500 stock price index. This variable's reported coefficient estimate in the DTA study differed dramatically from my originally reported results (see a_{13} in Table A1). The variable which should have been used was the period t-2 to t-1 return. The variable I misled DTA into using happened to be the period t-3 to t-2 return.

[4] At the time, I thought that, by putting the other items in my work and leisure queue ahead of responding more completely to the second replicator's request, I was being "optimally lazy." In retrospect and with *ex post* regret, "just plain lazy" is equally descriptive.

Coincidentally, interest in my 1983 paper produced a *third* replication effort by Skaggs (1984). Again, Skaggs' analysis was independent of and predated DTA's project. It was carried out as a preliminary to Skaggs' own extension of the literature on the real effects of unanticipated money growth. Skaggs estimated my money growth equation with independently collected data and obtained results very similar to those in my 1983 paper (even though his sample period differed slightly from mine).

Undoubtedly, Skaggs had been able to produce such results because, unlike DTA, he did not labor under the handicap of my direct assistance in data provision. However, in my opinion, DTA should have qualified their failure to replicate my published work by referencing Skaggs (1984). A citation of Skaggs' apparent success in independently reproducing my original results would have provided a balanced report on the probability that my 1983 results could have been replicated. I had previously called Skaggs' results to DTA's attention as part of my response to the second replicator's unsuccessful efforts.

APPENDIX

TABLE A1—ORIGINAL AND REPLICATION ATTEMPT RESULTS OF PROPERLY AND IMPROPERLY SPECIFIED REGRESSIONS

Coefficient	Merrick (1983)	Merrick (1986) on Merrick (1983)	DTA (1986) on Merrick (1983)	Merrick (1986) on DTA (1986)
a_0	.007 (.005)	.007 (.005)	.013 (2.41)	.013 (2.44)
a_1	.606 (.076)	.601 (.076)	.626 (6.85)	.622 (6.79)
a_2	−.054 (.092)	−.049 (.093)	−.039 (−0.38)	−.037 (−0.35)
a_3	.116 (.091)	.113 (.091)	.141 (1.40)	.139 (1.38)
a_4	−0.71 (.091)	−.069 (.091)	−.047 (−0.46)	−.046 (−0.46)
a_5	.097 (.078)	.090 (.079)	.099 (1.14)	.095 (1.09)
a_6	.075 (.070)	.080 (.069)	.065 (0.85)	.068 (0.89)
a_7	.0081 (.0035)	.0083 (.0034)	.010 (2.65)	.010 (2.69)
a_8	−.0048 (.0044)	−.0050 (.0044)	−.002 (−0.49)	−.002 (−0.49)
a_9	.0105 (.0065)	.0108 (.0066)	.011 (1.45)	0.11 (1.41)
a_{10}	−.0031 (.0041)	−.0033 (.0041)	−.004 (−0.93)	−.004 (−0.88)
a_{11}	−.273 (.087)	−.272 (.087)	−.380 (−4.06)	−.381 (−4.07)
a_{12}	.306 (.083)	.306 (.083)	.361 (3.90)	.362 (3.92)
a_{13}	.029 (.006)	.029 (.006)	−.0001 (−0.01)	.0001 (0.01)
DW	1.98	1.97	1.98	1.97

Notes: The Merrick (1983) column reports my original paper's results (standard errors, not *t*-ratios, in parentheses as in my original paper). Merrick (1986) on Merrick (1983) represents my recent attempt to reproduce these original results. This recent attempt was estimated using the TSP (Version 4.0B) package; Merrick (1983) used the IDA package. Both of these columns represent regressions with all data correctly aligned. DTA (1986) on Merrick (1983) is DTA's originally reported replication attempt using misaligned data (*t*-ratios in parentheses). Merrick (1986) on DTA (1986) represents my investigation of what went wrong in DTA's attempt. The a_{13} coefficient is crucially affected by the (unintended) misalignment.

REFERENCES

Dewald, William G., Thursby, Jerry G. and Anderson, Richard G., "Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*," *American Economic Review*, September 1986, *76*, 587–603.

Merrick, John J., Jr., "Financial Market Efficiency, the Decomposition of 'Anticipated' versus 'Unanticipated' Money Growth, and Further Tests of the Relation Between Money and Real Output," *The Journal of Money, Credit and Banking Project*: Comment," *American Economic Review*, May 1983, *15*, 221–32.

Skaggs, Neil T., "Real Effects of Monetary Shocks: Evidence from a Disaggregated Model," unpublished manuscript, Illinois State University, 1984.

Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*: Reply

By William G. Dewald, Jerry G. Thursby, and Richard G. Anderson*

The replication of published research is essential to progress in economic science. Proofs of mathematical theorems often may be replicated without contact with the original author; replication of published empirical research may be impossible even *with* the cooperation of the author. Our research suggests that the cause of this phenomenon is inadequate care and documentation by researchers, which is itself a response to a system of professional rewards which does not value replication. John Merrick's comments provide an example of the pitfalls in replication.

An important task of the *JMCB Project* was to ascertain the level of support and guidance an author would furnish to another researcher who was attempting replication. Merrick reviewed an earlier replication effort by Michael Zelina and offered constructive comments. Zelina had requested Merrick's data set from the *JMCB* office on January 30, 1984. On April 4, 1984, he wrote that he had independently checked and verified Merrick's variables DM and U against data published by Barro and Rush (1980) (who had furnished them to Merrick) but could not replicate Merrick's results. He said: "This leaves me to question the values of the remaining three variables $TBR1$, $FEDV$, and R_{t-1}. ...I wonder whether the values for R_{t-1}...are correct." (R_{t-1} is the variable mislabeled by Merrick.) On April 11, 1984, we forwarded these results to Merrick. Merrick replied on May 16, 1984: "I have begun to review my notes...to this paper...there are three possibilities...: (1) my original work is in serious error, (2) the data appendix I submitted is in error, and/or (3) your analyst either transcribed the data incorrectly or is otherwise in error. My current work schedule makes it impossible for me to devote my full attention to this matter until July." (Zelina was not our analyst and had no affiliation with the *JMCB Project*.) We did not receive further correspondence from Merrick.

Our replication efforts were conducted in the fall of 1984 and incorporated in a draft copy of the *JMCB Project Final Report to NSF* which we mailed to Merrick on January 15, 1985. In that report we wrote "...one researcher (Zelina, 1984) tried to replicate Merrick's (1983) results. Our results are very different from his [Zelina's] and very close to the ones reported by the author [Merrick]...Variable $[R^2_{t-2}]$ could be the source of the problems since our estimate of its coefficient has a different sign and a significantly different t-statistic than the ones reported in Table 1. This is the only major difference between our estimates and the author's estimates." We did not receive a reply.

It is unfortunate that we did not receive further communication from Merrick until after publication of our paper in this *Review* and scrutiny by his peers. In Merrick's comment we learn for the first time the reason for the difference in the estimates was because the supplied data set contained an error in the labels of one of the variables. As noted by Merrick in an earlier version of his comment, "One more exchange of information would have been helpful."

We published attempted replications of specific articles so as to demonstrate that the quality of record keeping practiced by empirical researchers often does not permit replication. Merrick's failure to understand the purpose of our work is illustrated by his references to Skaggs (1984) in both his letter of May 16, 1984, and his comment in this *Review*. Our purpose was *not* to ascertain the "correctness," accuracy, or robustness of published results. We did not suggest that results which we could not replicate are erroneous; indeed, we believe that well-documented erroneous results could be replicated perfectly. We sought to demonstrate that the materials supplied to us by some authors were inadequate to replicate. In addition, we noted in our article that our failure to replicate may be due to our errors. It is comforting to know that such is not the case for at least one of the papers which we failed to replicate.

---

*Bureau of Economic and Business Affairs, U.S. Department of State, Washington, DC 20520; Department of Economics, Krannert Building, Purdue University, West Lafayette, IN 47907; and Division of Monetary Affairs, Board of Governors, Federal Reserve System, Washington, DC 20551, respectively.

It is our belief that journals should publish the results of replication attempts—favorable or unfavorable—so as to improve the state of empirical economic research. But journals provide inappropriate incentives when they publish clarifying comments by authors who have failed to respond to requests for clarification prior to publication of negative results (even though one could hardly call our report on Merrick's work negative).

REFERENCES

Barro, Robert J. and Rush, Mark, "Unanticipated Money Growth and Economic Activity," in Stanley Fischer, ed., *Rational Expectations and Economic Policy*, Chicago: University of Chicago Press for the National Bureau of Economic Research, 1980, 23–49.

Dewald, William G., Thursby, Jerry G. and Anderson, Richard G., "Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*," *American Economic Review*, September 1986, *76*, 587–603.

Merrick, John J., Jr., "Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*: Comment," *American Economic Review*, December 1988, *78*, 1160–161.

Skaggs, Neil T., "Real Effects of Monetary Shocks: Evidence from a Disaggregated Model," unpublished manuscript, Illinois State University, 1984.

ERRATUM

Honesty in a Model of Strategic Information Transmission: Correction

By CAROLYN PITCHIK AND ANDREW SCHOTTER*

A number of typographical errors appeared in our paper published in the December 1987 issue of this *Review*. Our corrections appear below.

p. 1033 On line 2 above equation (1):
"probability $0 < \hat{p} < 1$, if the repair needed is"
should instead be
"probability $0 < 1 - \hat{p} < 1$, if the repair needed is"

p. 1033 On line 3 of footnote 1:
"$C_4 = \quad u(I|E) + (1-r)u(I|I)$; $E_1 =$"
should be
"$C_4 = ru(I|E) + (1-r)u(I|I)$; $E_1 =$"

p. 1034 On line 11 below A. Quality Control:
"decreases in r and that $\hat{q}$ is independent of"
should instead be
"increases in r and that $\hat{q}$ is independent of"

p. 1034 On line 7 below B. Price Controls:
"$u(E|I)$ remains the same. But then by"
should be
"$u(I|E)$ remains the same. But then by"

p. 1035 Left-hand column, equation in line 3 from the top:

$$\text{``}[r(u(E|E) - u|I|E))$$
$$-\hat{p} = (1-r)(u|I|I) - u|E|I))(1-s)]$$
$$\div [s(1-r)(u(I|I) - u|E|I))]\text{''}$$

should be

$$\hat{p} = [r(u(E|E) - u(I|E))$$
$$-(1-r)(u(I|I) - u(E|I))(1-s)]$$
$$\div [s(1-r)(u(I|I) - u(E|I))]$$

REFERENCE

Pitchik, Carolyn and Schotter, Andrew, "Honesty in a Model of Strategic Information Transmission," *American Economic Review*, December 1987, *77*, 1032–36.

---

*Departments of Economics, University of Toronto, Toronto, Ontario, Canada M5S 1A1; and New York University, New York, NY 10003, respectively.

Mansfield, E. 223
Marcet, A. 168
McCrate, E. 235
McElroy, M. B. 86
Melman, S. 55
Meltzer, A. H. 446
Mitchell, O. S. 66
Morton, J. S. 263
Nadiri, M. I. 429
Norwood, J. L. 284
Petersen, B. 200
Pfleiderer, P. 96
Phelps, E. S. 346
Pissarides, C. A. 363
Pollak, R. A. 242
Poterba, J. M. 413
Rebitzer, J. B. 389
Reynolds, B. L. 461
Riddell, T. 60
Rosenzweig, M. R. 245
Rosenberg, N. 229
Rowthorn, B. 194
Samuelson, W. 408
Sandler, T. 16
Sargent, T. J. 168
Scheiber, H. N . 140
Scherer F. M. 377
Shepherd, W. G. 50
Sims, Christopher A. 163
Snower, D. J. 38, 356
Solow, R. M. 378
Soper, J. C. 251
Steinmueller, W. E. 229
Stiglitz, J. E. 351
Summers, L. H. 182, 383
Summers, R. 467
Turner, J. A. 76
Viscusi, W. K. 300
Vishwanath, T. 110
Walstad, W. B. 251
Wilcox, S. P. 86
Yellen J. L. 44

◆

Segal, U. 233
Shaked, A. 837
Shapiro, C. 59
Shapiro, M. D. 595, 1067
Shugart, W. F., II 284
Siow, A. 108
Small, K. A. 560
Snower, D. 167
Solow, R. M. 307
Sonnenschein, H. 824
Spiegel, M. 824
Stahl, D. O., II 189
Stark, O. 1034
Stiglitz, J. 1046
Stockman A. C. 362
Summers, L. H. 273
Sussman, F. G. 255
Sutton, J. 837
Svensson, L. E. O. 375, 662
Tehranian, H. 1118
Thiel, S. E. 884
Thisse, J. 122
Thursby, J. G. 1162
Tollison, R. D. 284
Turnovsky, S. J. 341
Ursprung, H. W. 729
Van Huyck, J. B. 1088
Vives, X. 122
Wernerfelt, B. 246
West, K. D. 202
Winer, S. L. 701
Winston, C. 560
Wolff, E. N. 304, 1155
Young, L. 24
Zimmerman, M. B. 913

CONTRIBUTORS TO PAPERS AND PROCEEDINGS

Admati, A. R. 96
Akerlof, G. A. 44
Allen, B. 104
Alpern, S. 356
Auerbach, A. 211
Azariadis, C. 104
Barro, R. J. 32
Bassi, L. J. 91
Bauer, T. 452
Baumol, W. J. 257
Beller, A. H. 81
Bernanke, B. S. 435
Bernstein, J. I. 429
Blanchard, O. J. 173, 182
Blinder, A. S. 1, 435
Bluestone, B. 124, 377
Boskin, M. J. 401
Bowles, S. 145, 395
Boyer, R. 395
Brimmer, A. F. 151
Brunner, K. 446
Buchanan, J. M. 135
Buckles, S. 263
Cauley, J. 27
Chirinko, R. S. 206
Cockburn, I. 419
Coen, R. M. 188
Crandall, R. W. 323
Danzon, P. M. 305
Dickens, W. T. 129
Djajić, S. 341
Doescher, T. A. 76
Drazen, A. 369
Epstein, R. A. 311
Fazzari, S. 200
Fitoussi, J. 335
Freeman, R. B. 294
Friedman, B. M. 440
von Furstenberg, G. M. 178
Galbraith, J. K. 373
Gintis, H. 145
Glyn, A. 194
Gordon, D. M. 117
Gordon, R. J. 276
Graham, J. W. 81
Greenwald, B. 351
Griliches, Z. 419
Hall, R. E. 269
Harrison, B. 124
Heston, A. 467
Hickman, B. G. 188
Highsmith, R. J. 257
Hines, J. R., Jr. 211
Hubbard, R. G. 200
Im, E. I. 27
Jinglian, W. 461
Johnson, S. 408
Jorgenson, D. W. 217
Kahn, A. E. 316
Kane, E. J. 328
Kniesner, T. J. 86
Kotlikoff, L. J. 408
Krashevski, R. S. 289
Kravis, I. B. 474
Lang, K. 129
Lapan, H. E. 16
Leamer, E. E. 479
Le Cacheux, J. 335
Lee, D. R. 22
Leigh, W. A. 156
Levin, R. C. 424
Levine, P. B. 66
Lindbeck, A. 38
Lipsey, R. E. 474
Maasoumi, E. 178
Makarov, V. L. 457
Manchester, J. 70
Mankiw, N. G. 173

CONTRIBUTORS TO ARTICLES AND SHORTER PAPERS

Acs, Z. J. 678
Alesina, A. 796
Alm, J. 237
Anderson, R. G. 1162
Arnott, R. 1046
Audretsch, D. B. 678
Awerbuch, S. 581
Baldwin, R. E. 773
Ball, L. 999
Ballard, C. L. 1019
Basar, T. 341
Bateman, B. W. 1098
Baumol, W. J. 1155
Becker, G. S. 1
Bell, J. A. 282
Benjamini, Y. 281
Benjamini, Y. 281
Berger, M. C. 89
Bernheim, B. D. 1034
Binmore, K. 837
Blackorby, C. 691
Blomquist, G. C. 89
Borenstein, S. 913
Boyle, G. W. 24
Calvo, G. A. 647
Cecchetti, S. G. 999
Cheng, L. K. 746
Collier, I. L., Jr. 531, 613
Cooper, R. 138
Crawford, V. 485
Cropper, M. L. 255
Cutler, D. M. 1107
Danziger, L. 633
d'Aspremont, C. 1133
Deaton, A. 418
De Long, J. B. 273, 1138
Dewald, W. G. 1162
Dominguez, K. M. 595
Donaldson, D. 691
d'Orey, V. 341
Downs, T. 1118
Diba, B. T. 520
Dickson, V. 285
Drazen, A. 431
Eaton, J. 76
Eckstein, Z. 431
Economides, N. 108
Evans, L. 444
Fair, R. C. 595
Feldstein, M. 14
Field, E. B. 543
Frankel, J. A. 318
Friedman, D. 297
Garber, S. 444
Gianfrancesco, F. D. 277
Gibbons, R. 896
Gisser, M. 288
Greenwood, J. 402
Gregory, P. R. 531, 613
Grinols, E. L. 261
Grossman, G. M. 59
Grossman, H. I. 520, 1088
Helmberger, P. G. 46
Hendricks, K. 865
Hercowitz, Z. 402
Hernández, A. D. 362
Hettich, W. 701
Hillman, A. L. 729
Hochman, O. 290
Hoehn, J. P. 89
Hosios, A. J. 1046
Huberman, G. 471
Huffman, G. W. 402
Ishizawa, S. 794
Jacquemin, A. 1133
Johnson, W. R. 570
Jonung, L. 1080
Kaempfer, W. H. 786
Kahn, C. 471
Kimball, M. S. 224
King, S. R. 267
Kotlikoff, L. J. 662
Kroll, Y. 500
Kulatilaka, N. 574
Laidler, D. 1080
Lapan, H. E. 395
Layson, S. 1131
Lewis, T. R. 986
Lichtenberg, F. R. 550
Levy, H. 500
Lin, Y. J. 251
Lindbeck, A. 167
Loomes, G. 463
Lowenberg, A. D. 786
Luski, I. 290
McCormick, R. E. 284
MacDonald, G. M. 155
Marks, S. G. 574
Martin, R. E. 954
Matusz, S. 261
Merrick, J. J., Jr. 1160
Miranda, M. J. 46
Montgomery, C. A. 246
Moseley, F. 298
Murnighan, J. K. 806
Neary, P. 210
Neelin, J. 824
Papell, D. H. 531, 613
Pashigian, B. P. 936
Persson, T. 662
Pindyck, R. S. 969
Porter, R. H. 865
Rapoport, A. 500
Reinganum, J. F. 713
Rockett, K. 318
Ronn, E. I. 216
Roth, A. E. 806
Ruffin, R. J. 759
Sappington, D. E. M. 986
Schoumaker, F. 806

The Estimation and Measurement of Spillover Effects of R&D Investment

I. Cockburn and Z. Griliches: Industry Effects and Appropriability Measures in the Stock Market's Valuation of *R&D* and Patents 419

R. C. Levin: Appropriability, *R&D* Spending, and Technological Performance 424

J. I. Bernstein and M. I. Nadiri: Interindustry *R&D* Spillovers, Rates of Return, and Production in High-Tech Industries 429

Is It Money or Credit, or Both, or Neither?

B. S. Bernanke and A. S. Blinder: Credit, Money, and Aggregate Demand 435

B. M. Friedman: Monetary Policy Without Quantity Variables 440

K. Brunner and A. H. Meltzer: Money and Credit in the Monetary Transmission Process 446

Comparative Strategies for Economic Reform

T. Bauer: Economic Reforms Within and Beyond the State Sector 452

V. L. Makarov: On the Strategy for Implementing Economic Reform in the USSR 457

J. Wu and B. L. Reynolds: Choosing a Strategy for China's Economic Reform 461

Applications of International Comparisons of Prices and Quantities

A. Heston and R. Summers: What We Have Learned about Prices and Quantities from International Comparisons: 1987 467

I. B. Kravis and R. E. Lipsey: National Price Levels and the Prices of Tradables and Nontradables 474

E. E. Leamer: The Sensitivity of International Comparisons of Capital Stock Measures to Different "Real" Exchange Rates 479

◆

Technological Innovation and Productivity Change in Japan and the United States

D. W. Jorgenson: Productivity and Economic Growth in Japan and the United States 217

E. Mansfield: Industrial *R&D* in Japan and the United States: A Comparative Study 223

N. Rosenberg and W. E. Steinmueller: Why are Americans Such Poor Imitators? 229

Differing Theoretical Perspectives on the Household

E. McCrate: Gender Difference: The Role of Endogenous Preferences and Collective Action 235

R. A. Pollak: Tied Transfers and Paternalistic Preferences 240

M. R. Rosenzweig: Risk, Private Information, and the Family 245

High School Economics: Implications for College Instruction

W. B. Walstad and J. C. Soper: A Report Card on the Economic Literacy of U.S. High School Students 251

W. J. Baumol and R. J. Highsmith: Variables Affecting Success in Economic Education: Preliminary Findings from a New Data Base . . . 257

S. Buckles and J. S. Morton: The Effects of Advanced Placement on College Introductory Economics Courses 263

Inflation and Full Employment

R. E. Hall: Fluctuations in Equilibrium Unemployment 269

R. J. Gordon: The Role of Wages in the Inflation Process 276

The Measure and Character of American Unemployment

J. L. Norwood: The Measurement of Unemployment 284

R. S. Krashevski: What is So Natural About High Unemployment? 289

R. B. Freeman: Evaluating the European View that the United States Has No Unemployment Problem 294

Economic Aspects of Product Liability

W. K. Viscusi: Product Liability and Regulation: Establishing the Appropriate Institutional Division of Labor 300

P. M. Danzon: The Political Economy of Workers' Compensation: Lessons for Product Liability 305

R. A. Epstein: The Political Economy of Product Liability Reform 311

Surprises from Deregulation

A. E. Kahn: Surprises of Airline Deregulation . . . 316

R. W. Crandall: Surprises from Telephone Deregulation and the *AT&T* Divesture 323

E. J. Kane: Interaction of Financial and Regulatory Innovation 328

The Widespread Depression Overseas: American and Pacific Influences

J.-P Fitoussi and J. Le Cacheux: On Macroeconomic Implications of Price Setting in the Open Economy 335

S. Djajić: Exchange Rates, Wages, and the International Allocation of Capital 341

E. S. Phelps: A Working Model of Slump and Recovery from Disturbances to Capital-Goods Demand in an Open Nonmonetary Economy 346

Search Behavior in Labor and Product Markets

B. Greenwald and J. E. Stiglitz: Pareto Inefficiency of Market Economies: Search and Efficiency Wage Models 351

S. Alpern and D. J. Snower: "High-Low Search" in Product and Labor Markets 356

C. A. Pissarides: The Search Equilibrium Approach to Fluctuations in Employment 363

A. Drazen: Self-Fulfulling Optimism in an Trade-Friction Model of the Business Cycle 369

"The New Industrial State" After Twenty Years

J. K. Galbraith: Time and the New Industrial State 373

B. Bluestone: Discussion 377

R. M. Solow: Discussion 378

F. M. Scherer: Discussion 380

Efficiency Wages, Labor Relations, and Full Employment

L. H. Summers: Relative Wages, Efficiency Wages, and Keynesian Unemployment 383

J. B. Rebitzer: Unemployment, Labor Relations, and Unit Labor Costs 389

S. Bowles and R. Boyer: Labor Discipline and Aggregate Demand: A Macroeconomic Model . . 395

What Do We Know About Consumption and Saving, and What are the Implications for Fiscal Policy?

M. J. Boskin: Consumption, Saving, and Fiscal Policy 401

L. J. Kotlikoff, W. Samuelson, and S. Johnson: Consumption, Computation Mistakes, and Fiscal Policy 408

J. M. Poterba: Are Consumers Forward Looking? Evidence from Fiscal Experiments 413

CONTENTS OF THE PAPERS AND PROCEEDINGS

Richard T. Ely Lecture

A. S. Blinder: The Challenge of High Unemployment . . . 1

The Political Economy of Terrorism

H. E. Lapan and T. Sandler: To Bargain or Not To Bargain: That Is the Question . . . 16
D. R. Lee: Free Riding and Paid Riding in the Fight Against Terrorism . . . 22
J. Cauley and E. I. Im: Intervention Policy Analysis of Skyjackings and Other Terrorist Incidents . . . 27

The Natural Rate Theory Reconsidered

R. J. Barro: The Persistence of Unemployment . . . 32
A. Lindbeck and D. J. Snower: Long-Term Unemployment and Macroeconomic Policy . . . 38
G. A. Akerlof and J. L. Yellen: Fairness and Unemployment . . . 44

The Economics of the Arms Race

W. G. Shepherd: Self-Interest and National Security . . . 50
S. Melman: Economics Consequences of the Arms Race: The Second-Rate Economy . . . 55
T. Riddell: U.S. Military Power, the Terms of Trade, and the Profit Rate . . . 60

The Economics of the Aging of the Baby Boom

P. B. Levine and O. S. Mitchell: The Baby Boom's Legacy: Relative Wages in the Twenty-First Century . . . 66
J. Manchester: The Baby Boom, Housing, and Financial Flows . . . 70
T. A. Doescher and J. A. Turner: Social Security Benefits and the Baby-Boom Generation . . . 76

The Feminization of Poverty

A. H. Beller and J. W. Graham: Child Support Payments: Evidence from Repeated Cross Sections 81
T. J. Kniesner, M. B. McElroy, and S. P. Wilcox: Getting into Poverty Without a Husband, and Getting Out, With or Without . . . 86
L. J. Bassi: Poverty Among Women and Children: What Accounts for the Change? . . . 91

Markets for Information

A. R. Admati and P. Pfleiderer: Selling and Trading on Information in Financial Markets . . . 96
B. Allen and C. Azariadis: Informational Theories of Employment . . . 104
T. Vishwanath: Parallel Search and Information Gathering . . . 110

Challenging Some Conventional Wisdoms About the Labor Market

D. M. Gordon: The Un-Natural Rate of Unemployment: An Econometric Critique of the NAIRU Hypothesis . . . 117
B. Bluestone and B. Harrison: The Growth of Low-Wage Employment, 1963–86 . . . 124
W. T. Dickens and K. Lang: The Reemergence of Segmented Labor Market Theory . . . 129

Constitutional Convention Bicentennial

J. M. Buchanan: Contractarian Political Economy and Constitutional Interpretation . . . 135
H. N. Scheiber: Original Intent, History, and Doctrine: The Constitution and Economic Liberty . . . 140
S. Bowles and H. Gintis: Contested Exchange: Political Economy and Modern Economic Theory . . . 145

Issues in the Black Community

A. F. Brimmer: Income, Wealth, and Investment Behavior in the Black Community . . . 151
W. A. Leigh: The Social Preference for Fair Housing: During the Civil Rights Movement and Since . . . 156

Uncertainty in Macroeconomics

C. A. Sims: Uncertainty Across Models . . . 163
A. Marcet and T. J. Sargent: The Fate of Systems with "Adaptive" Expectations . . . 168
O. J. Blanchard and N. G. Mankiw: Consumption: Beyond Certainty Equivalence . . . 173
G. M. von Furstenberg and E. Maasoumi: Macroeconomic Implications of the Information Revolution . . . 178

Why Is Unemployment So High in Europe?

O. J. Blanchard and L. H. Summers: Beyond the Natural Rate Hypothesis . . . 182
R. M. Coen and B. G. Hickman: Is European Unemployment Classical or Keynesian? . . . 188
A. Glyn and B. Rowthorn: West European Unemployment: Corporatism and Structural Change . . . 194

Tax Policy and Investment: A Reconsideration

S. Fazzari, R. G. Hubbard, and B. Petersen: Investment, Financing Decisions, and Tax Policy . . . 200
R. S. Chirinko: Business Tax Policy, The Lucas Critique, and Lessons from the 1980's . . . 206
A. J. Auerbach and J. R. Hines, Jr.: Investment Tax Incentives and Frequent Tax Reforms . . . 211

T. Downs and H. Tehranian: Predicting Stock Price Responses to Tax Policy Changes 1118

S. Layson: Third-Degree Price Discrimination, Welfare and Profits: A Geometrical Analysis . . 1131

C. d'Aspremont and A. Jacquemin: Cooperative and Noncooperative *R&D* in Duopoly with Spillovers 1133

J. B. De Long: Productivity Growth, Convergence, and Welfare: Comment 1138

W. J. Baumol and E. N. Wolff: Reply 1155

J. J. Merrick, Jr.: Replication in Empirical Economics: *The Journal of Money, Credit and Banking Project*: Comment 1160

W. G. Dewald, J. G. Thursby, and R. G. Anderson: Reply 1162

◆

V. Crawford: Long-Term Relationships Governed by Short-Term Contracts 485

Y. Kroll, H. Levy, and A. Rapoport: Experimental Tests of the Separation Theorem and the Capital Asset Pricing Model 500

B. T. Diba and H. I. Grossman : Explosive Rational Bubbles in Stock Prices? 520

I. L. Collier, Jr. and D. H. Papell: About Two Marks: Refugees and the Exchange Rate Before the Berlin Wall 531

E. B. Field: The Relative Efficiency of Slavery Revisited: A Translog Production Function Approach 543

F. R. Lichtenberg: The Private *R&D* Investment Response to Federal Design and Technical Competitions 550

K. A. Small and C. Winston: Optimal Highway Durability 560

W. R. Johnson: Income Redistribution in a Federal System 570

N. Kulatilaka and S. G. Marks: The Strategic Value of Flexibility: Reducing the Ability to Compromise 574

S. Awerbuch: Accounting Rates of Return: Comment 581

K. M. Dominguez, R. C. Fair, and M. D. Shapiro: Forecasting the Depression: Harvard versus Yale 595

P. R. Gregory and I. L. Collier, Jr.: Unemployment in the Soviet Union: Evidence from the Soviet Interview Project 613

L. Danziger: Costs of Price Adjustment and the Welfare Economics of Inflation and Disinflation 633

G. A. Calvo: Servicing the Public Debt: The Role of Expectations 647

L. J. Kotlikoff, T. Persson, and L. E. O. Svensson: Social Contracts as Assets: A Possible Solution to the Time-Consistency Problem 662

Z. J. Acs and D. B. Audretsch: Innovation in Large and Small Firms: An Empirical Analysis 678

C. Blackorby and D. Donaldson: Cash Versus Kind, Self-Selection, and Efficient Transfers . . 691

W. Hettich and S. L. Winer: Economic and Political Foundations of Tax Structure 701

J. F. Reinganum : Plea Bargaining and Prosecutorial Discretion 713

A. L. Hillman and H. W. Ursprung: Domestic Politics, Foreign Interests, and International Trade Policy 729

L. K. Cheng: Assisting Domestic Industries under International Oligopoly: The Relevance of the Nature of Competition to Optimal Policies . . . 746

R. J. Ruffin: The Missing Link: The Ricardian Approach to the Factor Endowments Theory of Trade 759

R. E. Baldwin: Hysteresis in Import Prices: The Beachhead Effect 773

W. H. Kaempfer and A. D. Lowenberg: The Theory of International Economic Sanctions: A Public Choice Approach 786

S. Ishizawa: Increasing Returns, Public Inputs, and International Trade 794

A. Alesina: Credibility and Policy Convergence in a Two-Party System with Rational Voters . . . 796

A. E. Roth, J. K. Murnighan, and F. Schoumaker: The Deadline Effect in Bargaining: Some Experimental Evidence 806

J. Neelin, H. Sonnenschein, and M. Speigel: A Further Test of Noncooperative Bargaining Theory: Comment 824

K. Binmore, A. Shaked, and J. Sutton: Reply 837

K. Hendricks and R. H. Porter: An Empirical Study of an Auction with Asymmetric Information 865

S. E. Thiel: Some Evidence on the Winner's Curse 884

R. Gibbons: Learning in Equilibrium Models of Arbitration 896

S. Borenstein and M. B. Zimmerman: Market Incentives for Safe Commercial Airline Opertion 913

B. P. Pashigian: Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing . . 936

R. E. Martin: Franchising and Risk Management 954

R. S. Pindyck: Irreversible Investment, Capacity Choice, and the Value of the Firm 969

T. R. Lewis and D. E. M. Sappington: Regulating a Monopolist with Unknown Demand 986

L. Ball and S. G. Cecchetti: Imperfect Information and Staggered Price Setting 999

C. L. Ballard: The Marginal Efficiency Cost of Redistribution 1019

B. D. Bernheim and O. Stark: Altruism Within the Family Reconsidered: Do Nice Guys Finish Last? 1034

R. Arnott, A. J. Hosios, and J. Stiglitz: Implicit Contracts, Labor Mobility, and Unemployment 1046

M. D. Shapiro: The Stabilization of the U. S. Economy: Evidence from the Stock Market . . 1067

L. Jonung and D. Laidler: Are Perceptions of Inflation Rational? Some Evidence from Sweden 1080

H. I. Grossman and J. B. Van Huyck: Sovereign Debt as a Contingent Claim: Excusable Default, Repudiation, and Reputation 1088

B. W. Bateman: G. E. Moore and J. M. Keynes: A Missing Chapter in the History of the Expected Utility Model 1098

D. M. Cutler: Tax Reform and the Stock Market: An Asset Price Approach 1107

CONTENTS OF ARTICLES AND SHORTER PAPERS

G. S. Becker: Family Economics and Macro Behavior 1

M. Feldstein: The Effects of Fiscal Policies When Incomes Are Uncertain: A Contradiction to Ricardian Equivalence 14

G. W. Boyle and L. Young: Asset Prices, Commodity Prices, and Money: A General Equilibrium, Rational Expectations Model 24

M. J. Miranda and P. G. Helmberger: The Effects of Commodity Price Stabilization Programs 46

G. M. Grossman and C. Shapiro: Counterfeit-Product Trade 59

J. Eaton: Foreign-Owned Land 76

G. C. Blomquist, M. C. Berger, and J. P. Hoehn: New Estimates of Quality of Life in Urban Areas 89

N. Economides and A. Siow: The Division of Markets Is Limited by the Extent of Liquidity 108

J. Thisse and X. Vives: On the Strategic Choice of Spatial Price Policy 122

R. Cooper: Will Share Contracts Increase Economic Welfare? 138

G. M. MacDonald: The Economics of Rising Stars 155

A. Lindbeck and D. Snower: Cooperation, Harassment, and Involuntary Unemployment: An Insider-Outsider Approach 167

D. O. Stahl II: Bertrand Competition for Inputs and Walrasian Outcomes 189

K. D. West: On the Interpretation of Near Random-Walk Behavior in GNP 202

P. Neary: Determinants of the Equilibrium Real Exchange Rate 210

E. I. Ronn: Nonadditive Preferences and the Marginal Propensity to Consume 216

M. S. Kimball: Farmer's Cooperatives as Behavior Toward Risk 224

U. Segal: Does the Preference Reversal Phenomenon Necessarily Contradict the Independence Axiom? 233

J. Alm: Uncertain Tax Policies, Individual Behavior, and Welfare 237

B. Wernerfelt and C. A. Montgomery: Tobin's *q* and the Importance of Focus in Firm Performance 246

Y. J. Lin: Oligopoly and Vertical Integration: Note 251

M. L. Cropper and F. G. Sussman: Families and the Economics of Risks to Life 255

E. L. Grinols and S. J. Matusz: Some Welfare Implications of Job Mobility in General Equilibrium 261

S. R. King: Is Increased Price Flexibility Stabilizing? Comment 267

J. B. De Long and L. H. Summers: Reply 273

F. D. Gianfrancesco: The Choice Among Medical Insurance Plans: Comment 277

Y. Benjamini and Y. Benjamini: Reply 281

J. A. Bell: The Disinterest in Deregulation: Comment 282

R. E. McCormick, W. F. Shughart II, and R. D. Tollison: Reply 284

V. Dickson: Price Leadership and Welfare Losses in U.S. Manufacturing: Comment 285

M. Gisser: Reply 288

O. Hochman and I. Luski: Advertising and Economic Welfare: Comment 290

D. Friedman: Diamonds Are a Government's Best Friend: Burden-Free Taxes on Goods Valued for Their Values: Comment 297

F. Moseley: The Rate of Surplus Value, the Organic Composition, and the General Rate of Profit in the U.S. Economy, 1947–67: A Critique and Update of Wolff's Estimates 298

E. N. Wolff: Reply 304

R. M. Solow: Growth Theory and After 307

J. A. Frankel and K. Rockett: International Macroeconomic Policy Coordination When Policymakers Do Not Agree on the True Model 318

S. J. Turnovsky, T. Baser, and V. D'Orey: Dynamic Strategic Monetary Policies and Coordination in Interdependent Economies 341

A. C. Stockman and A. Hernández D.: Exchange Controls, Capital Controls, and International Financial Markets 362

L. E. O. Svensson: Trade in Risky Assets 375

H. E. Lapan: The Optimal Tariff, Production Lags, and Time Consistency 395

J. Greenwood, Z. Hercowitz, and G. W. Huffman: Investment, Capacity Utilization, and the Real Business Cycle 402

A. Deaton: Quality, Quantity, and Spatial Variation of Price 418

A. Drazen and Z. Eckstein: On the Organization of Rural Markets and the Process of Economic Development 431

L. Evans and S. Garber: Public-Utility Regulators Are Only Human: A Positive Theory of Rational Constraints 444

G. Loomes: When Actions Speak Louder Than Prospects 463

G. Huberman and C. Kahn: Limited Contract Enforcement and Strategic Renegotiation 471

Copyright 1988

All Rights Reserved

AMERICAN ECONOMIC ASSOCIATION

THE AMERICAN ECONOMIC REVIEW

VOLUME LXXVIII

BOARD OF EDITORS

GEORGE A. AKERLOF
JAMES E. ANDERSON
CLAUDIA D. GOLDIN
JO ANNA GRAY
GEORGE E. JOHNSON
KENNETH L. JUDD
JOHN F. KENNAN
MAURICE OBSTFELD
EDGAR O. OLSEN
ROBERT H. PORTER
JOHN G. RILEY
RICHARD ROLL
ALVIN E. ROTH
DAVID E. M. SAPPINGTON
KENNETH J. SINGLETON
ROBERT S. SMITH
BARBARA J. SPENCER
RICHARD TRESCH
LESLIE YOUNG

EDITOR

ORLEY ASHENFELTER

CO-EDITORS

ROBERT H. HAVEMAN

BENNETT T. McCALLUM

HAL R. VARIAN

THE AMERICAN ECONOMIC ASSOCIATION

Executive Office: Nashville, Tennessee

Editorial Office: 209 Nassau Street, Princeton, NJ 08542-4607

American Economic Association

Summer Minority Program

at Temple University

June 2, 1989–July 28, 1989

A New Economics Program with Exciting Features

An Attractive Stipend–plus Books,
Room, Board, Tuition and Travel are
Available to Qualified Students.

For More Information Contact:

DEPARTMENT OF ECONOMICS
RITTER HALL ANNEX
TEMPLE UNIVERSITY
PHILADELPHIA, PA 19122

(215-787-8887)

NORTH-HOLLAND

ECONOMICS JOURNALS

Carnegie Rochester Conference Series on Public Policy
1989: Vols. 30 & 31 in 2 issues

Economic Forecasts
1989: Volume 6 in 12 issues

Economics Letters
1989: Vols. 29–31 in 12 issues

European Economic Review
1989: Volume 33 in 9 issues

European Journal of Population
1989: Volume 5 in 4 issues

Information Economics and Policy
1989: Volume 4 in 4 issues

International Journal of Forecasting
1989: Volume 5 in 4 issues

International Journal of Industrial Organization
1989: Volume 7 in 4 issues

Japan and the World Economy NEW
1988/89: Volume 1 in 4 issues

Journal of Accounting and Economics
1989: Volume 11 in 4 issues

Journal of Banking and Finance
1989: Volume 13 in 6 issues

Journal of Development Economics
1989: Vols. 30 & 31 in 6 issues

Journal of Econometrics
(Including Annals of Applied Econometrics)
1989: Vols. 40–42 in 9 issues

Journal of Economic Behavior and Organization
1989: Vols. 11 & 12 in 6 issues

Journal of Economic Dynamics and Control
1989: Volume 13 in 4 issues

Journal of Economic Psychology
1989: Volume 10 in 4 issues

Journal of Financial Economics
1989: Vols. 23–25 in 6 issues

Journal of Health Economics
1989: Volume 8 in 4 issues

Journal of International Economics
1989: Vols. 26 & 27 in 8 issues

Journal of Mathematical Economics
1989: Volume 18 in 4 issues

Journal of Monetary Economics
1989: Vols. 23 & 24 in 6 issues

Journal of Public Economics
1989: Vols. 38–40 in 9 issues

Regional Science and Urban Economics
1989: Volume 19 in 4 issues

Resources and Energy
1989: Volume 11 in 4 issues

Studies in Banking and Finance
1989: Vols. 7 & 8 in 2 issues

The Review of Economics and Statistics
1989: Volume 71 in 4 issues

For more information or a free specimen copy please write to:

North-Holland
(An Imprint of Elsevier Science Publishers)

In the U.S.A. and Canada:
Elsevier Science Publishers
Attn: Journal Information Center
52 Vanderbilt Avenue
New York, NY, 10017, U.S.A.

In al other countries:
Elsevier Science Publishers
Attn: C. Fennes
P.O. Box 1991
1000 BZ Amsterdam, The Netherlands.

NH/ECON/JNL/6041d

AMERICAN LIVING STANDARDS

Threats and Challenges

ROBERT E. LITAN, ROBERT Z. LAWRENCE, and CHARLES L. SCHULTZE, editors

A central problem in the U.S. economy has been, and will continue to be, the slowdown in the improvement of living standards. In ***American Living Standards*** economists explain the causes and consequences of this slow down and examine what policy measures hold out the best hope for reversing it.

Cloth $29.95/paper $10.95

CRITICAL CHOICES

BARRY P. BOSWORTH, STEPHEN HESS,
ROBERT Z. LAWRENCE, ROBERT E. LITAN, THOMAS E. MANN,
CHARLES L. SCHULTZE, and JOHN D. STEINBRUNER

Critical Choices draws on the best research and analysis of Brookings to offer the next administration a framework for handling tough policy issues: reducing the deficit, limiting nuclear weapons, competing in the world economy. It outlines dynamic new policies and devises strategies for going beyond traditional political alliances to build a coherent consensus for action.

paper $6.95

CAN AMERICA AFFORD TO GROW OLD?

Financing Social Security

HENRY J. AARON, BARRY P. BOSWORTH, and GARY BURTLESS

Will taxes, along with existing surpluses, be sufficient to pay for benefits when the baby-boom generation retires? Economists Aaron, Bosworth, and Burtless examine the effects of rising social security costs and the measures to deal with them.

paper $7.95

CHANGING THE RULES

Technological Change, International Competition, and Regulation in Communications

ROBERT W. CRANDALL and KENNETH FLAMM

Competition is replacing regulation as a governing force in telecommunications. ***Changing the Rules*** brings together industry professionals and scholars to discuss the far-reaching implications of change within the industry and the current shifts in public policy.

Cloth $32.95/paper $12.95

BROOKINGS

Exam copies $3.00; request on letterhead

1775 Massachusetts Avenue NW
Washington, D.C. 20036 (202) 797-6258 *Visa/Mastercard*

HBJ Proudly Presents . . .

THE MODERN BESTSELLER

ECONOMICS
Principles and Policy

Fourth Edition

WILLIAM J. BAUMOL and ALAN S. BLINDER

Hardcover and two-volume paperbound, 1988

New for 1989

MICROECONOMICS
Theory and Applications
FRED R. GLAHE
DWIGHT R. LEE
Hardcover, 1989
Just Published!

INTRODUCTORY ECONOMETRICS WITH APPLICATIONS
RAMU RAMANATHAN
Hardcover, 1989
Just Published!

COMPARING ECONOMIC SYSTEMS
A Political Economy Approach
Second Edition
ANDREW ZIMBALIST,
HOWARD JAY SHERMAN,
and STUART S. BROWN
Hardcover, 1989
Just Published!

Also Available

LABOR AND THE ECONOMY
Second Edition
HOWARD M. WACHTEL
Hardcover, 1988

INTERMEDIATE MICROECONOMICS
ROBERT W. CLOWER,
PHILLIP E. GRAVES,
and ROBERT SEXTON
Hardcover, 1988

THE EVOLUTION OF ECONOMIC THOUGHT
Fourth Edition
JACOB OSER
STANLEY L. BRUE
Hardcover, 1988

HBJ

HARCOURT BRACE JOVANOVICH, Inc.
College Sales Office
7555 Caldwell Avenue, Chicago, IL 60648
(312) 647-8822

INTERNATIONAL
ECONOMIC JOURNAL

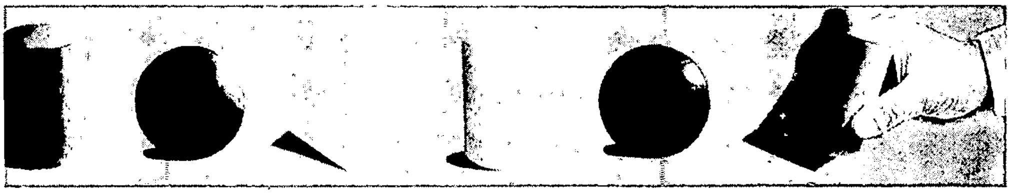
VOLUME 2 AUTUMN 1988 NUMBER 3

Estimating Investment and Savings Functions for Developing Countries: with an Application to Latin America — Nathaniel H. Leff and Kazuo Sato
Strategic Trade Policies, Differentiated Duopoly and Intra-Industry Trade — K. C. Fung
Foreign Borrowing as Dutch Disease: A Quantitative Analysis for the Philippines — Romeo M. Bautista
Factor Demand in Pakistan's Manufacturing Sector — Ashfaque H. Khan
Quantity Expectations and Unemployment in Non-Walrasian Economics — Chen-Min Hsu
Estimates of Capital Stock and Capital/Output Coefficients by Industries for Korea: 1953-86 — Hak-Kil Pyo

International Economic Journal is published quarterly in English by the Korea International Economic Association. The personal subscription price of a volume (which includes postage) is $30.00 per year. The institutional subscription price is $40.00 per year. Cheques should be made payable to the International Economic Journal and sent directly to the Managing Editor. (Volume 1 is available.)

MANUSCRIPTS SUBMITTED *(in triplicate) for publication from Asia should be sent to the Managing Editor, Professor Wontack Hong, Department of International Economics, College of Social Sciences, Seoul University, Seoul 151-742, Korea. Manuscripts from countries outside Asia should be sent to the Co-Editor, Professor Young Chin Kim, Department of Economics, Northern Illinois University, Dekalb, Illinois 60115, U.S.A.*

This spring, a lot of micro students will exceed expectations—thanks to Eaton and Eaton's MICROECONOMICS



Eaton and Eaton is the intermediate micro text based on the notion that learning economics is doing economics. That notion is working.

Published just last year, Eaton and Eaton is already in use at a long list of schools across the country. How does a new text become so popular so quickly? With exceptionally clear writing and careful development of ideas to help students grasp important material. Find out for yourself. Eaton and Eaton is worth a close look.

"A well thought-out book. Especially attractive is the approach that microeconomics involves building models and solving problems."
—Peter Huang, *Tulane University*

1988, hardbound, 471 pages, ISBN 0-7167-1818-9 • With complete supplementary package

W. H. Freeman and Company
The book publishing arm of *Scientific American*
41 Madison Avenue, New York, NY 10010

Scott, Foresman Economics Texts *The Perfect Competition*

Byrns and Stone's Total Teaching Package—Why Accept Less?

Economics

Fourth Edition
Ralph T. Byrns, University of Colorado, Boulder
Gerald W. Stone, Metropolitan State College

This best-selling text is the cornerstone of a total teaching package completely author-produced that includes a versatile collection of print supplements and computer software (MACROSIM macroeconomic simulation, RAIDERS microeconomic tutorial, StudyCalc interactive graphing program and spreadsheet, and computerized Test Banks). The most "reader-friendly" introductory text on the market, it represents the authors' commitment to provide instructors with quality products that make teaching economics far easier and more enjoyable.

Macro Features

- A new chapter on Federal Budget Deficits and the National Debt includes material on the link between budget deficits and trade deficits, as well as on international comparisons of deficits and savings rates.
- International dimensions are covered throughout, including differences in the velocity of money, the Dance of the Dollar (exchange rate volatility), the New Protectionism, the International Debt Crisis, and Japan's Industrial Policy.
- Expanded, extensive use of summary tables that do such things as delineate factors influencing Aggregate Demand or Aggregate Supply and break down the policy differences between the various macro schools.

Micro Features

- A new chapter explores the unique microeconomic problems of agriculture (Ch. 34).
- Law and Economics sections contain economic analysis of legal cases on such topics as insider information, seniority and the value of the marginal product of labor, and compliance with restrictions on growing crops.
- Expanded coverage of welfare reform proposals, the impact of foreign companies on the competitive position of U.S. firms, and changes in the workforce and productivity.

Available Now, *©1989, in single-volume and split editions, with Instructor's Manual, Study Guide, Graphing Tutorial, Study-Calc interactive spreadsheet, RAIDERS microeconomic simulation software, MACROSIM II macroeconomic simulation software, Test Banks, EXAM-SCIENCE, EXAM-WORD, DIPLOMA electronic classroom management system, and Transparencies.*

Expanded coverage of financial institutions and international issues!

The Economics of Money, Banking, and Financial Markets

Second Edition

Frederic S. Mishkin, Columbia University

This best-selling money and banking text is now revised and updated for even greater effectiveness. The first text to use simple economic concepts as a unifying framework for exploring the complex and rapidly changing financial environment continues its real-world approach, consistently emphasizing the interaction of economic theory with empirical data, and presenting simple models that encourage understanding rather than memorization. Mishkin's crisp, to-the-point prose makes even difficult ideas accessible to students. The text offers maximum flexibility for various course structures.

New to the Second Edition

- Expanded coverage of financial markets and institutions
- Increased coverage and integration of international issues
- Thoroughly updated coverage, including discussions of the October 1987 stock market crash; recent Federal Reserve policy; the crisis facing banks, S&Ls and the FSLIC; and junk bonds and program trading
- New interactive software that allows students to practice the art of economic analysis

Available Now, *©1989, 768 pp., hardbnd., with Instructor's Manual/Test Bank, Study Guide/Workbook, Computerized Test Bank, Transparency Masters, and "The FED Game and Financial Market Analysis" interactive software.*

An Ideal Supplement

Current Readings on Money, Banking, and Financial Markets

1988 Edition

James A. Wilcox

This inexpensive reader is updated annually to help students stay current with the rapid changes in the field. *©1988, 370 pp., softbnd.*

Scott, Foresman and Company

Scott, Foresman Economics Texts
The Perfect Competition

Basics at Their Best!

Microeconomic Theory and Applications

Third Edition

Edgar K. Browning / Jacquelene M. Browning
Texas A&M University

The best-selling text in this field offers lucid, in-depth coverage of basic topics. Four separate applications chapters provide additional applications, giving students an unparalleled understanding of the theoretical topics and their ramifications.

Third Edition highlights

- Five new extensive applications, including intertemporal choice and human capital; excise taxation; the NCAA cartel; sugar price support program; and the Consumer Price Index
- 100 new end-of-chapter problems
- A new end-of-text math appendix that is keyed to the main text for easy, optional use
- Newly revised Test Bank with over 200 additional questions

Available Now, *©1989, 672 pp., hardbnd., with Instructor's Manual/Test Bank; Study Guide/Workbook; Interactive Software.*

Finally, a one-term principles text that offers real depth

Contemporary Economics

J.R. Kearl, Brigham Young University

Kearl's new one-semester principles text emphasizes the depth, not the breadth, of micro and macro topics found in the big two-semester texts. A strong micro core is developed before the essentials of modern macro are covered.

Features include:

- International economics integrated early into the text
- Strong presentation of the micro foundations of macroeconomics
- More real-world applications than in competing texts
- Extensive and superb end-of-chapter problems
- A strong public policy theme throughout
- Lucid, nontechnical coverage of information economics and uncertainty

Available Now, *©1989, 575 pp., softbnd., with Instructor's Manual, Test Bank, and Study Guide.*

Brief
Nontechnical
Traditional Topic Order
Applications-Oriented
Modern

▲ **Basic Economics**
Paul R. Gregory / Roy J. Ruffin
University of Houston

From the authors of the best-selling *Principles of Economics* comes a new text that more fully emphasizes their acclaimed pedagogy and writing style. Briefer and with a more traditional topic order (e.g., Keynesian economics appears before the money chapters), this is the first introductory text to identify and display all key economic ideas in a wide margin format. Full-color layout is combined with abundant photographs and a single text column per page to create a less intimidating learning format. **Available Now,** *©1989, in single-volume and split editions, with Instructor's Manual, Study Guide, Micro and Macro disks developed exclusively for this text, Test Bank, Overhead Transparencies, and DIPLOMA electronic classroom management system.*

Scott, Foresman and Company

Scott, Foresman Economics Texts *The Perfect Competition*

New for 1988-89

▲ **The Antitrust Revolution**
John E. Kwoka, Jr.
Lawrence J. White

Thirteen non-technical studies of recent antitrust cases, written by the economists who participated in them, offer students a clear illustration of the microeconomic analysis used in today's legal cases. **Available Now,** *©1989, 330 pp., softbnd.*

▲ **The Economics of the Employment Relationship**
Robert J. Flanagan
Lawrence M. Kahn
Robert S. Smith
Ronald G. Ehrenberg

An integrated treatment of the economic, legal, and organizational aspects of the employment relationship in both union and neoclassical labor models, with a comparative systems thread throughout. **Available January 1989,** *620 pp., hardbnd., with "The Competitive Edge" software simulation program.*

▲ **Urban Economics, 4/e**
Edwin S. Mills / Bruce Hamilton

Ideal for advanced undergraduate courses, this one-term text uses economic theory to analyze cities, urban issues, and problems. The Fourth Edition features stronger coverage of the historical evolution of cities, an economic analysis of urban education, and much more. **Available Now,** *©1989, 420 pp., hardbnd.*

▲ **State and Local Public Finance**
Institutions, Theory, Policy
Ronald C. Fisher

Assuming only a micro principles background, this text uses microeconomic theory to analyze thoroughly the practice and problems encountered in state and local government taxation and expenditure. **Available Now,** *©1989, 550 pp., hardbnd.*

▲ **Comparative Economic Systems**
Heinz Kohler

This innovative text explores alternative criteria underlying any economic system, and focuses on Marxist theory, socialist systems, the Soviet economy, Yugoslavia, China, the LDC model, and modern Western welfare states. SYSTEMS-1, an IBM disk with programs on input/output, linear programming, cash flow, other investment criteria, and problem sets, is free with each new text. **Available February 1989,** *610 pp., illustrated, hardbnd.*

▲ **Money and Capital Markets**
Tim S. Campbell

Analytical microeconomics-based text with the strongest coverage yet of interest rate determination, portfolio theory, and asset liability management of financial intermediaries in a competitive deregulated environment. *©1988, 608 pp., hardbnd.*

Recently Published

Economics, 2/e
Robert B. Ekelund / Robert D. Tollison

Clear, balanced, and complete coverage of modern economics is integrated into the traditional core. More than any other basic book, it focuses on public policy issues and how economics affects people's everyday lives. New macro chapter, "Fiscal Policy, Taxation, and the Debt," plus new material on the effects of the 1986 tax reform. *©1988, available in single-volume and split editions, with complete supplement package.*

Principles of Economics, 3/e
Roy J. Ruffin / Paul R. Gregory

New open economy theme throughout the micro section; new Keynesian work on expectations, as well as Neoclassical updates; integration of multiyear contracting, layoff unemployment, and contract markets as sources of inflexibility. The only basic text with a chapter on information economics—updated with new applications. Over 100 new boxed applications throughout. *©1988, available in single-volume and split editions, with complete supplement package, including both software "games" and tutorials for IBM PC.*

Introductory Economics
Marilu McCarty

Economics Tool Kit, an integrative computer software package, comes with each copy at no extra cost; various text passages and problem sections are keyed to the software. Unique macro chapter on the role of financial intermediation, including financial futures. Strong public policy flavor (e.g., in health economics) and non-technical, applications-oriented presentation. *©1988, available in single-volume and split editions, with complete supplement program.*

Scott, Foresman and Company

Scott, Foresman
Economics Texts
The Perfect Competition

Recently Published

Economics: Principles of Political Economy, 3/e
Daniel R. Fusfeld

An extensive update of the only text to offer a systematic introduction to economics with a historical, institutionalist orientation. New early chapter on the international economy. *©1988, 864 pp., hardbnd., with complete supplement program.*

International Economics: Theory and Policy
Paul R. Krugman / Maurice Obstfeld

An ideal combination of up-to-date theory and real-world policy issues, making it the most current book available. The author's systematic presentation of theory enables students to analyze key policy issues. *©1988, 650 pp., hardbnd., with Instructor's Manual.*

Modern Macroeconomics
Geoffrey Woglom

Innovative new text blends Keynes, rational expectations, and money. The only text to offer separate chapters on market-clearing and nonmarket-clearing rational expectations models, the term structure of interest rates, and labor microfoundations. *©1988, 780 pp., hardbnd.*

Modern Price Theory
Roy J. Ruffin

Intermediate-level text with worked-out algebraic problem sets covers such high-interest topics as information economics, Shepherd's lemma, screening, moral hazard, and adverse selection, in addition to strong traditional coverage. *©1988, 500 pp., hardbnd.*

Law and Economics
Robert Cooter / Thomas Ulen

Strong economic analysis of property rights, contracts, torts, and crime, with a wealth of modern examples. Early chapter on legal thinking and economic thinking and how they differ. *©1988, 630 pp., hardbnd.*

Microeconomics with Calculus
Brian R. Binger
Elizabeth Hoffman

For honors intermediate, masters level, or background for doctoral course. Two chapters on uncertainty; new welfare economics; experimental economics applications; duality theory. ©*1988, 610 pp., hardbnd.*

Modern Labor Economics, 3/e
Ronald G. Ehrenberg
Robert S. Smith

Updated and expanded, with new work on efficiency wages, mobility migration, and two rewritten macro chapters. New 400-page Student Workbook loaded with problem sets. ©*1988, 608 pp., hardbnd., with Instructor's Manual and Student Workbook.*

Statistics for Business and Economics, 2/e
Heinz Kohler

The most applications-oriented statistics text is revised to include a vast number of problem sets. With low-cost DAYSAL-2 program and problem-solving disks. ©*1988, 775 pp., hardbnd., with Instructor's Manual and Student Workbook.*

Environmental and Natural Resource Economics, 2/e
Thomas H. Tietenberg

Three new chapters—on water scarcity, forest resources, and regional and global air pollutants—highlight this revision. ©*1988, 521 pp., hardbnd.*

Understanding Futures Markets, 2/e
Robert W. Kolb

This inexpensive paperback explores the various types of futures instruments, the market, and the role of regulation. Special emphasis on financial futures. ©*1988, 430 pp., softbnd.*

Principles of Finance
Robert W. Kolb

Modern text progresses from money and capital markets to portfolio theory, then to the firm and its financing and investing decisions. Separate intuitive chapters on options, futures, and hedging strategies. Strong coverage of traditional topics such as risk, return, and capital budgeting. ©*1988, 840 pp., hardbnd., with full supplement program*

▲ **For further information write**
Meredith Hellestrae,
Dept. SA-AER
1900 East Lake Avenue
Glenview, Illinois 60025

Scott, Foresman and Company

Blackwell

Lectures on Advanced Econometric Theory
DENIS SARGAN
Edited by MEGHNAD DESAI
$39.95 208pp 0-631-14956-2

The New Classical Macroeconomics
KEVIN D. HOOVER
"critical writing of the very highest order. The breadth of coverage is also very impressive." – Robert E. Lucas
$49.95 304pp 0-631-14605-9

Economic History and the Modern Economist
WILLIAM N. PARKER
"represents some of the best writing – and writers – the field has to offer." – *Choice*
Now Available in Paperback!
$12.95 (pb) 120pp 0-631-15585-6

On Ethics and Economics
AMARTYA SEN
$12.95 132pp 0-631-15494-9

Economists and the Economy
The Evolution of Economic Ideas 1600 to the Present Day
ROGER BACKHOUSE
$49.95 240pp 0-631-15558-9

Multinationals and the European Community
Edited by JOHN H. DUNNING and PETER ROBSON
Journal of Common Market Studies
$19.95 (pb) 168pp 0-631-16329-8

The Emergence of the Welfare States
DOUGLAS E. ASHFORD
Now Available in Paperback!
$18.95 (pb) 352pp 0-631-16023-X

Do Socialist Economies Work?
The Soviet and East-European Experience
MICHAEL BLEANEY
$49.95 256pp 0-631-15306-3

Growth, Development and Welfare
AJIT K. DASGUPTA
$49.95 224pp 0-631-14399-8

Technology, Education and Productivity
Essays in Applied Econometrics
ZVI GRILICHES
$39.95 288pp 0-631-15614-3

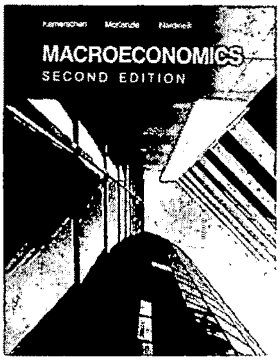
Surveys in Public Sector Economics
Edited by PAUL G. HARE
$49.95 288pp 0-631-15686-0

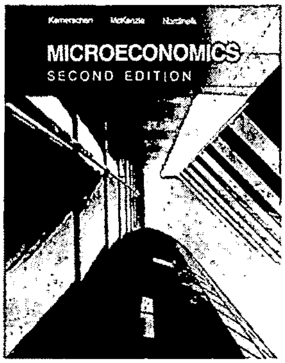
Innovation, Technology and Finance
Edited by ARNOLD HEERTJE
$49.95 216pp 0-631-15952-5

Pre-Industrial Economic Growth
Social Organization and Technological Progress in Europe
GUNNAR PERSSON
$45.00 208pp 0-631-14963-5

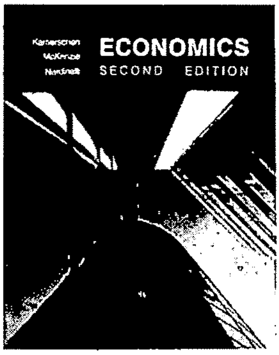
☎ **Toll-Free Ordering:**
1-800-638-3030

Basil Blackwell
432 Park Avenue South
New York, NY 10016





VALUE
ADDED



Economics, Second Edition
David Kamerschen, University of Georgia
Richard McKenzie, Clemson University
Clark Nardinelli, Clemson University

Complete Hardcover Edition • about 1,024 pages
Macroeconomics, Second Edition • paperback about 576 pages
Microeconomics, Second Edition • paperback about 608 pages

Instructor's Manual • Test Bank • PC Test Bank Plus • Call-in Test Service • Transparencies Quadrant I: Color Graphics Software • Study Guide • MicroStudy: Computerized Study Guide PC Economics: Interactive Graphing Software with Workbook • GPA: Grade Performance Analyzer • Just published

The new Kamerschen/McKenzie/Nardinelli author team brings valuable expertise in macro, micro, and history of economics to the Second Edition. Retaining the First Edition's flexible organization and in-depth coverage, *Economics, Second Edition* is the first introduction to incorporate fully integrated pedagogy that helps students learn concepts quickly and more efficiently. Chapter-opening Key Questions, which focus on the chapter's central points, are repeated in the margin next to the appropriate text area. At the end of each chapter, Key Questions are summarized for review.

Economics, Second Edition also features more Perspectives boxes and new Dialogue boxes that present contrasting views of current economic issues. New to the support package is PC Test Bank Plus with graphing capabilities.

For adoption consideration, request an examination package from your regional Houghton Mifflin office.

Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
101 Campus Dr., Princeton, NJ 08540

The Economic Forecast Is BASIC

NEW

THE GREAT U-TURN
Corporate Restructuring
and the Polarizing of America
BENNETT HARRISON & BARRY BLUESTONE
$19.95

LIABILITY
The Legal Revolution
and Its Consequences
PETER W. HUBER
$19.95

MEMOIRS OF AN UNREGULATED ECONOMIST
GEORGE J. STIGLER
Sloan Foundation Science Series
$17.95

RENDEZVOUS WITH REALITY
The American Economy after Reagan
MURRAY WEIDENBAUM
$19.95

SOCIAL SCIENCE IN GOVERNMENT
Uses and Misuses
RICHARD P. NATHAN
$18.95

NEW TEXT EDITIONS

OPENECONOMYMACROECONOMICS
Second Edition
RUDIGER DORNBUSCH
$24.95

PRINCIPLES OF MONEY, BANKING, AND FINANCIAL MARKETS
Sixth Edition
LAWRENCE S. RITTER & WILLIAM L. SILBER
$36.95

At your bookstore or direct
from the publisher at 1-800-638-3030.
Major credit cards accepted.
Libraries and institutions
call 1-800-242-7737.

NEW IN PAPERBACK

BAILOUT
An Insider's Account
of Bank Failures and Rescues
IRVINE S. SPRAGUE
$9.95

MANUFACTURING MATTERS
The Myth of the Post-Industrial Economy
STEPHEN S. COHEN & JOHN ZYSMAN
$9.95

THE TRANSFORMATION OF AMERICAN INDUSTRIAL RELATIONS
THOMAS KOCHAN, HARRY KATZ, & ROBERT MCKERSIE
$11.95

UNHEARD VOICES
Labor and Economic Policy
in a Competitive World
RAY MARSHALL
$9.95

COMING IN SPRING '89

BALANCING ACTS
Technology, Finance,
and the American Future
JAMES K. GALBRAITH
$21.95

FLUCTUATING FORTUNES
The Political Power of Business
in America
DAVID VOGEL
$20.95

THE NEW COMPETITORS
How Foreign Investors Are Changing
the U.S. Economy
NORMAN J. GLICKMAN
& DOUGLAS P. WOODWARD
$19.95

WRONG NUMBER
The Breakup of AT&T
ALAN STONE
$19.95

BASIC BOOKS **BASIC BOOKS, INC.** 10 E.53rd St., New York 10022

INTERNATIONAL MONETARY FUND
STAFF PAPERS

Staff Papers, a quarterly journal, contains studies by the Fund's staff on international monetary and financial problems. The December 1988 issue includes:

The Balance of Trade, Terms of Trade, and the Real Exchange Rate: An Intertemporal Optimizing Framework. . . . . by Jonathan Ostry

The Effects of Currency Substitution on the Response of the Current Account to Supply Shocks. . . . . by Carlos Vegh

Urban and Rural Household Savings in China. . . . . by Yingyi Qian

Debt Relief and Adjustment Incentives: A Theoretical Exploration. . . . . . by W. Max Corden

Multilateral Negotiations for Rescheduling Developing Country Debt: A Bargaining-Theoretic Framework. . . . . by Jeremy Bulow and Kenneth Rogoff

Financial Deregulation, Demand for Money, and Monetary Policy in Australia. . . . . .by P.A.V.B. Swamy and George S. Tavlas

The Sustainability and Optimality of Government Debt. . . . . .by Howell H. Zee

Subscription rates: US$18.00 a volume, US$5.00 for a single issue.
Special rates for university libraries, faculty, and students: US$9.00 a volume.
Available from: Publications Services • Box E-434
International Monetary Fund • 700 19th Street, N.W.
Washington, D.C. 20431, U.S.A. • Telephone (202) 623-7430

FEDERAL RESERVE BANK OF NEW YORK
RESEARCH FUNCTION

Research positions available for experienced economists and new Ph.Ds. in macroeconomics, monetary theory and policy, finance, international economics and applied econometrics. Temporary and permanent appointments available. Salaries commensurate with experience and level of achievement. Equal Opportunity Employer. For more information contact:

A. Steven Englander
Recruiting Officer
Federal Reserve Bank of New York
33 Liberty Street, Room 913
New York, New York 10045
(212) 720-5635

ECRS DIRECTORY OF UNPUBLISHED RESEARCH - BUSINESS & ECONOMICS

BASE VOLUME

A UNIQUE AND INVALUABLE SOURCE OF RESEARCH, DATA, INFORMATION, AND IDEAS.

Indexes over 43,000 unpublished works such as working / discussion / technical / seminar papers and many other types of research in all areas of business and economics, from the USA and abroad.

The DIRECTORY indexes: (a) author, (b) title, (c) affiliation, (d) I.D.#, (e) year, (f) type (g) subject category. It is organized by subject.

Dimensions - 2,000 pages; 143 subject areas; 650 institutional sources; up to 20 years; 160 work types; USA and world-wide; 9"x12"x3.

143 subject areas on a cross-section of theory, institutions and applications in micro/macro economics, policy, accounting, finance, marketing, O.R., engineering, law, applied math, stats, credit, econometrics, regulation, R&D, industry/country studies, monetary/fiscal topics, political economy, history, agriculture, bibliographies, computers, education, demography, insurance, women, poverty, pension, socialism, minorities, environment, energy, and many, many more (request complete list on your letter-head).

The number of institutions and individuals registering research on an ever increasing range of topics with ECRS is steadily rising. SUPPLEMENTS, with some 10,000 additional items, will be made available to registered BASE VOLUME purchasers at a fraction of the current price. The BASE VOLUME will remain the only comprehensive cummulative source.

Price: $88, plus $11 for UPS and handling. **AEA member discount: 50%.** Payment must accompany orders. Make checks payable to "ECRS, INC."

Address orders and other correspondence to:
ECONOMIC CONSULTING & RESEARCH SYSTEMS - ECRS INC
200 Park Ave - P.O. Box 1278
New York, NY 10163-1278

Oxford

New!
An unrivaled resource for courses in economic history

A Concise Economic History of the World

From Paleolithic Times to the Present

RONDO CAMERON, *Emory University*

Grounded in modern economic theory but written in non-technical language accessible to non-specialists, this comprehensive history highlights events that have contributed to prosperity and material well-being—that is, to "development"—as well as those that have resulted in continued "underdevelopment" for large areas of the world. The first half of the book surveys world history from ancient times to the eighteenth century, outlining the origins of Western prosperity. The second half investigates in detail the industrialization and urbanization that transformed Europe and America in the nineteenth and twentieth centuries. Cameron places the contemporary world economy in broad historical perspective, making both the East-West rivalry and the North-South split more understandable. With its predominantly narrative style, *Economic Development in History* is invaluable for upper division and graduate courses in economic history and economic development.

March 1989 464 pp. illus. $29.95

Prices and publication dates are subject to change.
To request an examination copy, write on school letterhead giving full course information, including course name, level, expected enrollment, and your decision deadline, to:
College Humanities & Social Sciences Marketing Dept.

Oxford University Press
200 Madison Avenue • New York, NY 10016

Oxford

World Development Report 1988

THE WORLD BANK

"Basic source of data on LDC's and complementary reading to any of the leading textbooks in Economic Development."
—D.K. Zschock, *State University of New York at Stony Brook*

"A basic and essential summary of the views of the world's premier development institution, with the supporting data."
—John W. Thomas, *Harvard University* (on the 1987 Report)

Recognized as an indispensable resource for economists, policymakers, journalists, and professionals of all kinds, the *World Development Report* continues to survey and synthesize contemporary economic issues in developed and developing countries. The 1988 Report surveys opportunities and risks in adjustment and growth worldwide, with a special section on public finance in growth development, providing an up-to-date, informative analysis of recent trends with an abundance of supporting data. As always, the Report incorporates research from all over the world, analyzing the conditions that surround the growth process and offering insights on appropriate policy measures for the future.

(A World Bank Publication)

1988 320 pp.; charts, tables, graphs, and maps
paper $12.95 cloth $26.00

Again available in paperback—by the winner of the 1987 Nobel Prize in Economics

Growth Theory

An Exposition

ROBERT M. SOLOW, *Massachusetts Institute of Technology*

With American productivity growing at ever-smaller rates and international stock markets in spectacular turmoil, this distinguished book by one of the greatest living economists is more timely than ever. The paperback edition features the full text of Solow's Nobel Prize acceptance speech.
1988 142 pp.; 11 illus. paper $10.95

Survival and Change in the Third World

Edited by BEN CROW and MARY THORPE, both of the *Open University of Britain*

Following the development of Third World economies over a five-hundred-year period, this unique collection provides a much-needed interdisciplinary framework for further debate on development and underdevelopment in the Third World; examining not only industrial and agricultural production, but also the people involved in running Third-World economies.
1988 376 pp. paper $14.95 cloth $39.95

Dynamic Economic Analysis

MILTON HARRIS, *Graduate School of Business, The University of Chicago*

"Excellent compact survey."—Sudipto Bhattacharya, *University of Michigan.*
"An excellent text....Provides a clear, concise presentation of important dynamic economic modelling concepts."—Jamie Kruse, *University of Colorado*
1987 152 pp. $19.95
Solutions Manual free to instructors

Microeconomic Principles

FRANK A. COWELL, *The London School of Economics and Political Science*

This comprehensive textbook offers thorough coverage of the basic principles that underlie standard modern theoretical and applied economics. Among the topics covered are production, the firm, the consumer, households and workers, aggregation, equilibrium, efficiency, uncertainty, incentives, and welfare.
1986 422 pp. $26.00

The Economics of Uncertainty

C.J. McKENNA, *University College, Cardiff*

"Very appealing."—Jonas Prager, *New York University*. The first work of its kind, this non-technical introduction outlines broad issues and various techniques, and then considers their applications in such areas as insurance, quality uncertainty, signalling, and the stock market.
1986 139 pp. $19.95

Prices and publication dates are subject to change.
To request an examination copy, write on school letterhead giving full course information, including course name, level, expected enrollment, and your decision deadline, to:
College Humanities & Social Sciences Marketing Department

Oxford University Press
200 Madison Avenue • New York, NY 10016

COMPARATIVE WORLD DATA

A Statistical Handbook for Social Science

compiled by Georg Müller

Compiled at the Sociological Institute of the University of Zurich, *Comparative World Data* offers a wide range of social, economic, political, cultural, and demographic information covering 130 countries. For each country, 51 variables provide information ranging from per capita income and infant mortality rates to share of Nobel Prize winners and extent of civil liberties.

For more sophisticated analyses, the entire data base is available for purchase with the book on IBM-compatible diskettes, but the volume's format permits exploratory data analysis without use of a computer. A source of encyclopedic information on the world's countries as well as a device for making cross-national comparisons, testing hypotheses, and constructing theories, *Comparative World Data* provides an invaluable research tool.

$49.95 *hardcover*
$87.00 *hardcover with 3 5¼" or 2 3½" diskettes*

SUBSIDIZING INDUSTRIAL LOCATION

A Conceptual Framework with Application to Korea

Michael P. Murray

How can policymakers in developing countries best induce firms to locate at particular sites? Michael P. Murray develops the first rigorous conceptual framework for analyzing the economic efficiency of alternative policies for influencing firms' choices of location. Rooted in a theory of optimal subsidies, the framework relies on econometric and simulation models of firms' costs.

World Bank Occasional Papers, New Series, Number 3

$14.95 *paperback*

EXCHANGE RATE MISALIGNMENT IN DEVELOPING COUNTRIES

Sebastian Edwards

Sebastian Edwards identifies the factors determining the real exchange rate at equilibrium and examines the causes of misalignment under fixed, floating, and dual (or black market) exchange rate regimes. He stresses the relation between the misalignment rate and macroeconomic policies, describes how to measure the costs of misalignment, and evaluates strategies to realign real exchange rates.

World Bank Occasional Papers, New Series, Number 2

$9.95 *paperback*

THE JOHNS HOPKINS UNIVERSITY PRESS
701 West 40th Street, Suite 275, Baltimore, Maryland 21211

AMERICAN ECONOMIC ASSOCIATION

1989 ANNUAL MEMBERSHIP RATES

Membership includes:

—a subscription to *The American Economic Review* (quarterly) plus *Papers and Proceedings*, the *Journal of Economic Literature* (quarterly) and the *Journal of Economic Perspectives* (quarterly).

- Regular members with annual incomes of $30,000 or less $42.00
- Regular members with annual incomes above $30,000 but no more than $40,000 .... $50.40
- Regular members with annual incomes above $40,000 .... $58.80
- Junior members (available to registered students for three years only). Student status must be certified by your major professor or school registrar $21.00
- In Countries other than the U.S.A., Add $16.00 to cover postage.
- Family members (persons living at the same address as a regular member, additional memberships without subscription to the publications of the Association) $8.40

Please enter my membership for the following period:

☐ **Jan.-Dec.** ☐ **April-March** ☐ **July-June** ☐ **Oct.-Sept.**

First Name and Initial **Last Name** **Suffix**

Address Line 1

Address Line 2

City

State or Country **Zip/Postal Code**

MAJOR FIELDS (TWO ONLY)

LIST FIELDS WITH WHICH YOU CURRENTLY IDENTIFY. SELECT FIELD CODE FROM *JEL*, "Classification System for Books."

☐ ☐

Please type or print information above. Please pay with a check or money order payable in United States Dollars. Canadian and foreign payments must be in the form of a draft or check drawn on a United States bank payable in United States Dollars. Please note: It is the policy of the Association, not to refund membership payments.

Endorsed by (AEA member) ____________________

Below for Junior Members Only

I certify that the person named above is enrolled as a student at ____________________

Authorized Signature

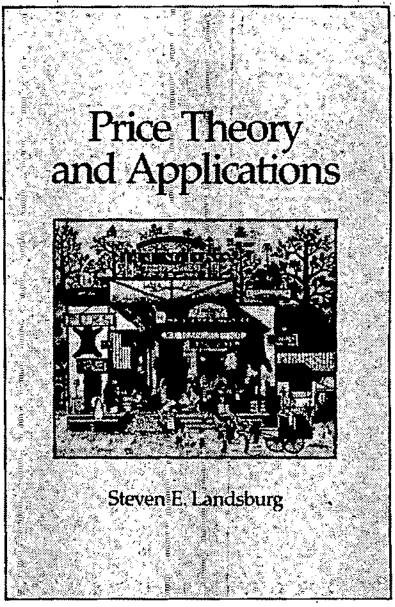
PLEASE SEND WITH PAYMENT TO:

AMERICAN ECONOMIC ASSOCIATION
1313 21ST AVENUE SOUTH, SUITE 809
NASHVILLE, TENNESSEE 37212-2786
U.S.A.

The Most Modern Applied Treatment of Intermediate Microeconomics Available!

Price Theory and Applications

Steven E. Landsburg,
University of Rochester



This major new contribution to the field of price theory will show students how consumers and business firms make decisions, and how these decisions interact in the marketplace. The text encourages students to develop and use economic intuition by submitting it to precise verbal and graphical analysis. Without relying on intimidating mathematical tools, PRICE THEORY AND APPLICATIONS exposes students to sophisticated economic reasoning, and includes examples and problems that will challenge even the most ambitious and talented student.

Comprehensive Ancillary Package Includes:

Instructor's Manual • Test Bank • Study Guide with GraphPac • Computerized Test Banks • Transparency Acetates • Computer Assisted Instruction • Landsburg's Exercises and Calculus Supplement

1989 656 pages hardcover ISBN 0-03-020589-1

THE DRYDEN PRESS
908 North Elm Street, Suite 101, Hinsdale, IL 60521

D61-8AER-F

University of Siena – C.N.R.
INTERNATIONAL SCHOOL OF ECONOMIC RESEARCH

announces a
Workshop on COMPETING EXPLANATIONS OF HIGH UNEMPLOYMENT
July 4-16, 1989, at the Certosa di Pontignano, Siena (Italy).

The International School of Economic Research was established in 1987 and has organized two previous workshops.

The first on "Alternative Approaches to Macroeconomics" was held in November of 1987 with the participation of the following lecturers:

J.-P. Benassy (Cepremap)
M. De Cecco (Roma)
D. Hendry (Oxford)
J. R. Hicks (Oxford)
A. Leijonhufvud (UCLA)
B. T. McCallum (Carnegie-Mellon)
A. Quadrio Curzio (Milano)
T. J. Sargent (Hoover Inst.)
J. Taylor (Stanford)
R. Day (Southern Cal.)
R. M. Goodwin (Siena)
D. Heymann (Buenos Aires)
D. Laidler (U.W. Ontario)
S. Lombardini (Torino)
E. S. Phelps (Columbia)
J. Sachs (Harvard)
J. E. Stiglitz (Princeton)
A. Vercelli (Siena)
M. Woodford (Chicago)

The School's second workshop on "Economics and Institutions" was held in July of 1988 with the participation of:

P. Bardhan (UC Berkeley)
R. W. Clower (South Carolina)
H. Demsetz (UCLA)
J. R. Hicks (Oxford)
A. Leijonhufvud (UCLA)
U. Pagano (Siena)
F. Romani (Roma)
A. Schotter (New York U.)
S. Bowles (U. Mass.)
M. De Cecco (Roma)
R. Heiner (Brigham Young)
B. Holmstrom (Yale)
D. North (U. Washington)
E. S. Phelps (Columbia)
R. Rowthorn (Cambridge)
M. Shubik (Yale)
H. Simon (Carnegie-Mellon)

Inquiries concerning the faculty for Workshop III can be directed to the Secretary of the School and will be answered as soon as arrangements have been completed. The Siena School workshops are intended for advanced graduate students and junior faculty members. The members of the faculty will be in residence and available to students in and out of lecture and debate sessions for the entire duration of the workshop. Enrollment will be limited to 30 fellows. Room and board will be provided by the School but, in principle, fellows will have to pay for their own travel. Applications, including a detailed curriculum vitae and two letters of reference, should arrive before May 1st, 1989, and be addressed to:

Barbara Engelmann
The Secretary International School of Economic Research
Dipartimento di Economia Politica, Università di Siena, Piazza S. Francesco, 53100 Siena (Italy), tel. (+577) 298-619

Available for Text and Reference Use

REGULATION AND MARKETS

Daniel F. Spulber

Regulation and Markets provides the up-to-date, integrated analysis of regulatory policies and the administrative process that is needed in today's field of regulation economics. The book takes a modern perspective, using the tools of industrial organization and game theory. It is the only unified treatment of the field and combines theoretical models with consideration of public policy issues in the areas of antitrust, price regulation, environmental regulation, product quality, and workplace safety.
February, 1989 712 pp. $45.00

55 Hayward Street
Cambridge, MA 02142

The MIT Press

Journal of Economic Development

Welcomes unsolicited manuscripts and subscriptions by Interested Parties

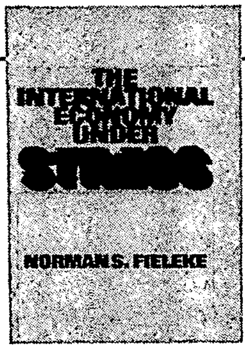
Managing Editor
IN KIE KIM
Chung-Ang University
Co-Editors:
PAN A. YOTOPOULOS
Stanford University
CHOONG YONG AHN
Chung-Ang University
Editorial Board:
DALE W. ADAMS
Ohio State University
RICHARD H. DAY
University of Southern California
MASANORI HASHIMOTO
Indiana University
LAWRENCE B. KRAUSE
UC., San Diego
YOUNG JOON LEE
Pusan National University
RONALD I. McKINNON
Stanford University
F. I. NIXSON
University of Manchester
JEFFREY B. NUGENT
University of Southern California
ATTIAT F. OTT
Clark University
SEUNG PARK
Chung-Ang University
HAK YONG RHE
Korea University
BYUNG NAK SONG
Seoul National University
U. TUN WAI
International Monetary Fund
ALAN RUFUS WATERS
California State University, Fresno
JANG H. YOO
Virginia Commonwealth University
SUK BUM YOON
Yonsei University

The Journal of Economic Development (JED) is published biannually in June and December since 1976 by The Economic Research Institute of Chung-Ang University, Seoul, Korea. The JED contains articles on a wide range of issues concerning economic development, especially in LDCs with a particular attention to NICs members.

- The JED emphasizes empirical and policy aspects of economic development. Articles in the JED are annotated by the American Economic Association in the Journal of Economic Literature.
- The JED has served as an important communications vehicle among scholars and professionals in universities, governments, international agencies, and development research institutes of developing countries with deep interests in economic development in general as well as specific policy issues confronting them.
- The JED welcomes unsolicited manuscripts, which will be considered for publication by the Editorial Board. The Editorial Board consists of leading scholars in economic development. Three copies of the manuscript should be accompanied by an abstract of no more than 100 words and a brief curriculum vitae including the author's academic career. All submissions should be typewritten, double-spaced in English with footnotes, references, figures, tables and any other illustrative materials on separate sheets.
- For subscription to the Journal (US$26 for individuals and US$46 for institutions per year), send a check or money order payable to The Economic Research Institute, Chung-Ang University, Heuksuk-Dong, Dongjak-Ku, Seoul 156-756, Korea.

Manuscripts and Editorial Correspondence should be addressed to:

CHOONG YONG AHN
Editor, The Journal of Economic Development
The Economic Research Institute
Chung-Ang University
Seoul 156-756, Korea



The International Economy Under Stress

Norman S. Fieleke

"Norman Fieleke offers a remarkably comprehensive yet brief account of the world economy during nearly two turbulent decades. He covers the main developments — from oil shocks to U.S. deficits — in an integrated and illuminating fashion; and his treatment has the outstanding merit of being quantitative, thus putting into proper perspective the apparently contradictory qualitative arguments and interpretations that figure so heavily in public discussion."

— Richard N. Cooper
Maurits C. Boas Professor of International Economics
Harvard University

"Fieleke's book admirably combines rigor of economic analysis with a refreshing clarity of style and organization. Throughout these chapters the treatment is balanced, sensible, and fair-minded. Students of the international economy could ask for no more instructive survey of contemporary issues."

— Benjamin J. Cohen
William L. Clayton Professor of International Economic Affairs
Fletcher School of Law and Diplomacy
Tufts University

September 1988	272 pages
0-88730-224-6	cloth, $29.95
0-88730-340-4	paper, $14.95

The U.S. Business Corporation

An Institution in Transition

John R. Meyer and
James Gustafson, Editors
An American Academy of Arts and Sciences Book

"In the complacent era of the 1970s, the question was how to make the U.S. business corporation more open and democratic. This volume still asks "for whom does the corporation toil," but it also asks how to make it toil more effectively. Distinguished contributors from many walks of life help to answer these questions in a lively, readable fashion. Give a copy to each of your directors. The epilogue alone is worth the price."

— Louis Lowenstein
Professor of Finance and Law
Columbia University
Author of *What's Wrong with Wall Street*

November 1988	264 pages
0-88730-354-4	cloth, $34.95

YES! Please send me:
____**The International Economy Under Stress**
Cloth (6611990) $29.95
Paper (6613392) $14.95
____**The U.S. Business Corporation** (6613145) $34.95
☐ Payment enclosed ☐ Bill me
Charge my ☐ MC ☐ VISA ☐ AMEX
Card no.__________Exp. date______
Signature__________
Send to: __________
____
__________Zip______

BALLINGER
PUBLISHING COMPANY

Order Department
P.O. Box 1630
Hagerstown, MD 21741

My state sales tax $______
Postage/handling ($2.75/order)* $______
***Prepaid orders are postage free!**
TOTAL $______

Prices subject to change. All orders subject to credit approval. U.S. funds only. If you order by phone, tell the operator your order code is
AAER1288

The AMEX® Bank Review
AWARDS

The Editors of The AMEX Bank Review are pleased to announce the winners of the 1988 Essay Competition in International Economics and Financial Markets, in memory of Robert Marjolin. A former adviser to the Review, Professor Robert Marjolin was the first Secretary-General of the OECD and former Vice President of the European Commission, and was a leading architect of the European Community.

The first three prize winning essays are now published in the Review's Special Paper series, available from the Editors. All the essays will be published in early 1989, jointly with Oxford University Press.

FIRST PRIZE $15,000
William Emmott
The Economist, London.
"The Limits to Japanese Power"

SECOND PRIZE $5,000
Allen Lenz
Chemical Manufacturers Association, Washington D.C.
"Slimming the US Trade and Current Account Deficits"

THIRD PRIZE $2,500
Mark Mullins
London School of Economics
"Meltdown Monday or Meltdown Money: Causes of the Stock Market Crash"

YOUNG ECONOMIST PRIZE $2,000
Menzie Chinn
University of California, Berkeley, USA.
"A Tale of Two Markets: Capital Market Integrity and the Determination of Exchange Rates"

MID-EAST/AFRICA PRIZE $2,000
Phedon Nicolaides
Royal Institute for International Affairs, London
"Globalization of Services and Developing Countries"

LATIN AMERICA PRIZE $2,000
Alvaro Zini
University of Sao Paulo, Brazil
"Brazil at the Crossroads: Foreign Debt and Fiscal Exhaustion"

ASIA PRIZE $2,000
Bejoy Das Gupta
Christ Church, Oxford University
"Determinants of Exports from Developing Countries"

SPECIAL MERIT AWARDS

Special Merit Awards ($1,000) were given to: Roy Allen, St Mary College of California; Nicholas Butler, British Petroleum Exploration; David Cobham, University of St. Andrews (U.K.); Joshua Greene, International Monetary Fund; Albert Hahn, EcoPlan International (Paris); Christopher Huhne, The Guardian (London); David Meerschwam, Harvard Business School; Anthony Saunders, New York University and Federal Reserve Bank of Philadelphia; Zvi Schloss, J. Rothschild Investment Management Ltd.; and Brian Tew, Loughborough University (U.K.).

The Essays were judged by the Review Editors and a special Award Committee: Lord Roll of Ipsden (President, S. G. Warburg Group plc); Professor Peter Kenen, (Princeton University); Bruce MacLaury, (President, The Brookings Institution); Bahram Nowzad, (Chief Editor, International Monetary Fund); Kevin Pakenham, (Chief Executive, John Govett & Co. Limited); Richard O'Brien, (Chief Economist, American Express Bank Ltd).

Details of the 1989 Essay Competition (to be announced in early 1989) will be obtainable from the Editors at American Express Bank Ltd., 60 Buckingham Palace Road, London SW1W ORR

AMERICAN
EXPRESS
BANK

AEA sponsored Group Life Insurance for you and your family—at attractive rates!

The AEA Group Life Insurance Plan can help provide valuable supplementary protection—at attractive rates—for eligible members and their dependents.

Because AEA participates in a large Insurance Trust which includes other scientific and technical organizations, the low cost may be even further reduced by premium credits. In the past nine years, insured members received credits on their April 1 semiannual payment notices averaging 40% of their annual premium contributions. (These credits are based on the amount paid during the previous policy year ending September 30.) Of course future premium credits, and their amounts, cannot be promised or guaranteed.

Now may be a good time for you to re-evaluate your present coverage and look into AEA Life Insurance. Just fill out and return the coupon for more details at no obligation.

Administrator, AEA Group Insurance Program
1255 23rd Street, N.W.
Washington, D.C. 20037

J-8

Please send me more information about the AEA Life Insurance Plan.

Name ______________________________ Age ________

Address ______________________________

City ____________________ State ________ Zip ________

Or—call today Toll-Free 800-424-9883
(Washington, DC area, call 296-8030)

THE ECONOMICS INSTITUTE

Gateway to Successful
Master's and Doctoral Degree Studies
in Economics, Agricultural Economics, Business and Administration

Since its establishment in 1958 under the sponsorship of the American Economic Association, the Institute has provided specialized preparatory training and orientation to over 6,000 students from over 120 countries en route to over 320 universities in the United States and other English-speaking countries.

PROGRAM FEATURES

- Six to forty-six week individualized programs.
- Intensive and comprehensive review or supplementary preparatory training in:
  English, Computer Usage, Economic Theory, Mathematics, Statistics, Accounting, Management, Finance and Marketing.
- Standardized test preparation (TOEFL, GRE, GMAT).
- Orientation to U.S. campus and community life.
- University placement assistance services.
- Certificate and Diploma Programs for short-term professional trainees.

END RESULTS for

- **Associated universities:** Improved admissions procedures and a source of additional high-quality foreign students.
- **Foreign students:** Greater accessibility to U.S. universities.
- **Sponsoring organizations:** Reduced total costs.
- **And for all three:** Better grade performance records in degree programs and more rapid completion of degree requirements.

ORGANIZATION

AEA Policy and Advisory Board:
Edwin S. Mills, Northwestern University, Chairman
Lance E. Davis, California Institute of Technology
Koichi Hamada, Yale University
Joseph Havlicek, Ohio State University
Teh-wei Hu, University of California, Berkeley
Samuel A. Morley, Vanderbilt University
Ray Marshall, University of Texas at Austin
Stefan H. Robock, Columbia University

Director
Wyn F. Owen, University of Colorado

The Economics Institute, 1030 13th Street, Boulder, Colorado, 80302, USA;
Telephone: (303)492-3000; Telex: 450385 ECONINST BDR; FAX: (303)492-3006

South Africa Inc.

The Oppenheimer Empire

David Pallister, Sarah Stewart, and Ian Lepper

The first in-depth investigation of the powerful Oppenheimer family of South Africa, whose powerful cartel controls much of the world's diamond, gold and platinum supply. The authors reveal for the first time the full dimensions of the Oppenheimers' web of secret holdings, the extent of their economic and political power, and the fallacies behind their carefully cultivated anti-apartheid stance.

"**A highly valuable contribution** to our understanding of the role of big business in South Africa and its influence over the global economy." —Leonard Thompson $25.00

The Evolution of Policy Behind Taiwan's Development Success

K. T. Li

with introductory essays by Gustav Ranis and John C. H. Fei

A high-level minister who was one of the main architects of organizational and policy change in Taiwan analyzes the success of Taiwan's economic development over the past forty years. K. T. Li's account provides insight into the institutional underpinnings of a successful development experience and draws conclusions about the features of Taiwan's development that could be instructive for other developing countries. $25.00

NEW PAPERBOUNDS

In Whose Interest?

International Banking and American Foreign Policy

Benjamin J. Cohen

The first detailed study of the relationship between the American banking community and government policymakers.

"Could well become **a classic commentary on international finance today.** . . . Provides a **clear and convincing** diagnosis of the main ways that finance and foreign policy affect one another." —Edward L. Morse, *The New York Times Book Review* $12.95

A Council on Foreign Relations Book

Biotechnology: The University-Industrial Complex

Martin Kenney

"Kenney's work is the first major effort to provide a detailed analysis of the birth of the new industrial field of biotechnology and its impact on universities. . . . Kenney's book **abounds in rich description and valuable conjectures.** It also provides important insights into the structural and institutional aspects of the biotechnological revolution."
—Sheldon Krimsky, *American Scientist*
$12.95

Common Decency

Domestic Policies After Reagan

Alvin L. Schorr

with a chapter by James P. Comer

Planner, educator, and activist Alvin Schorr presents an incisive critique of America's social policies under the Reagan administration and offers clear, reasonable, and workable proposals for improving policies in income distribution, housing, health, and education.

"**A provocative and stimulating** discussion."
—Alan Shank, *Perspective* $9.95

Leadership at the Fed

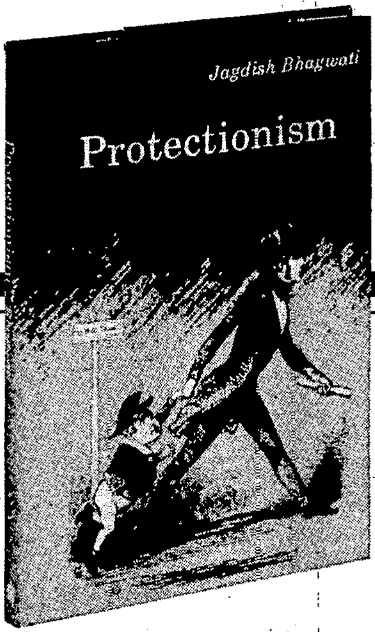
Donald F. Kettl

A political history of the Federal Reserve Board and of the men who guided its growth into the national and international force it is today.

"In this **extensively researched** book, Donald F. Kettl. . . . confirms beyond a doubt that since its inception in 1913 the Federal Reserve has become a paramount power in the making of economic policy. . . . **Authoritative.**"
—Lawrence A. Veit, *The New York Times Book Review* $11.95

Yale University Press
Dept. 846
92A Yale Station
New Haven, CT 06520

The Economic Side



"A 'tour de force'!"
—Paul A. Samuelson

Protectionism
Jagdish Bhagwati
"Bhagwati's lectures on protection . . . combine modern economic analysis, profound insight into the political pressures working for and against a retreat from the 40-year march to freer trade, and his accustomed wit and brilliance."—Charles P. Kindleberger, MIT
Illustrated $16.95

Exchange Rates and Inflation
Rudiger Dornbusch
Brings together for the first time, the writings that form the basis of Rudiger Dornbusch's contribution to open economy macroeconomics and exchange rate theory.
$27.50

Privatization
An Economic Analysis
John Vickers and George Yarrow
Offers a comprehensive analysis of the British privatization program, along with insights into recent policies on privatization, competition, and regulation in that country.
$39.95

Inflation Stabilization
The Experience of Israel, Argentina, Brazil, Bolivia, and Mexico
Michael Bruno, Guido Di Tella, Rudiger Dornbusch, and Stanley Fischer
Provides a critical analysis of the disinflation programs introduced in four countries in 1985–86, and discusses the possibility of such a program in Mexico.
$27.50

The Theory of Industrial Organization

Jean Tirole

The first primary text to treat the new industrial organization at the advanced-undergraduate and graduate level. Rigorously analytical and filled with exercises coded to indicate level of difficulty, it provides a unified and modern treatment of the field.
$35.00

Now Available in Paperback

The Evolution of Central Banks

Charles Goodhart

Many economists have challenged the value of central banks, but Charles Goodhart provides an authoritative and original approach in applying the considerable insights of history to concerns of current policymaking.
$11.95 paper ($22.50 cloth)

Policies for Prosperity

Essays in a Keynesian Mode

James Tobin

edited by Peter M. Jackson

In these accessible and timely essays, James Tobin shows how Keynesian economics offer corrective treatment for the economic ailments we have faced in the Ford, Carter, and Reagan presidencies.
$14.95 paper ($35.00 cloth)

The International Monetary System

Essays in World Economics

Richard N. Cooper

Unless exchange rate uncertainty is reduced, Cooper foresees growing restrictions on trade and capital movement. In this incisive book, he offers a solution to this crucial problem.
$9.95 paper ($32.50 cloth)

NBER Macroeconomics Annual 1988

edited by Stanley Fischer

$14.95 paper ($30.00 cloth)

55 Hayward Street
Cambridge, MA 02142

The MIT Press

JAPAN, INC.
Introduction to Japanese Economics (The Comic Book)
SHOTARO ISHINOMORI
With an Introduction by PETER DUUS

New in cloth and paper—

"At last: a book that makes Japan's economy seem simple."
—*Newsweek*

"Now Americans seeking clues to how Japan views itself, the United States and economic issues generally can see for themselves."—*New York Times*

"Behind the bright, jokey cover, and woven into the plot's salacious diversions to strip shows in the Ginza and blackmail at an S&M brothel, is serious writing on political economics."
—*San Francisco Examiner*

$25.00 cloth, $10.95 paper

IMMIGRANT ENTREPRENEURS
Koreans in Los Angeles, 1965-1982
IVAN LIGHT and EDNA BONACICH

This book offers the most comprehensive case study ever completed of the causes and consequences of immigrant business ownership. Analyzing both the short-run and the long-run causes of Korean entrepreneurship, the authors explain why the Koreans could find, acquire, and operate small business firms more easily than could native born residents.
$45.00

OUR DAILY BREAD
The Peasant Question and Family Farming in the Colombian Andes
NOLA REINHARDT

Our Daily Bread makes a valuable contribution to development studies by tackling head-on the current debate over the persistence of the family farm. Reinhardt examines the dynamics of family farm agriculture and analyzes the impact of national policies on its competitive position in a southwest Andean community.
$35.00

New in paper—

PROSPERITY AND UPHEAVAL
The World Economy 1945-1980
HERMAN VAN der WEE

"A monumental book, in which the author surveys masterfully, yet with a light touch, the whole of the world's economic development from the end of the war to c. 1980."—*Economic History Review*

"A useful survey of the theories of economic growth and development. It will also be valuable in broadening the perspective of Americans, since Van der Wee. . . has a European orientation."
—*Library Journal*

$14.95 paper

At bookstores or call toll-free 800-822-6657. Visa and MasterCard only.

University of California Press • Berkeley 94720

Economics has more to offer than free market economics, mathematical abstrusity, and the Marxist paradigm. . .

If you DISAGREE, do not read further!

If you AGREE, consider teaching a course in your department that uses:

Economics as Social Science: An Evolutionary Approach

Wendell Gordon *The University of Texas, Austin*
and
John Adams *The University of Maryland, College Park*

Economics as Social Science is a new and modern exposition of the ideas of evolutionary (American institutional) economics. The book is a useful and stimulating introduction to the methods and theory of contemporary institutional economics. Gordon and Adams contrast the evolutionary approach with standard microeconomics and macroeconomics, and with Marxism. As a text, Economics as Social Science may be used in courses in non-orthodox and heterodox economics, or economic ideas.

If you are UNDECIDED, answer the following true-false questions: 1. The institutions, technology, and values of society do not change, as orthodox economic theory usually assumes. (FALSE) The world in which we live is a dynamic, changing place. 2. People are driven exclusively by rational gain considerations (profit and utility maximization), as conventional economists assume. (FALSE) Managers and consumers act on the basis of complex motives which they themselves do not fully understand. 3. Firms usually operate under conditions of perfect competition, as standard price theory alleges. (FALSE) Oligopolies are widely prevalent and it is difficult to identify even one major competitive industry in the modern American economy. 4. Economics increasingly resembles formal mathematics rather than a science of society. (TRUE) Orthodox economics uses intricate and opaque geometry, algebra, and calculus to try to prove its recondite arguments. 5. Economics is purely "positive" and "value free." (FALSE) Economics is a social science dealing with human behavior and has an obligation to explain how people's values arise and change.

About Examination Copies:

If you AGREED with the initial proposition, or if you answered correctly three or more of the true-false questions, then you are invited to request an examination copy of:

Economics as Social Science: An Evolutionary Approach
by
Wendell Gordon and John Adams

Cloth $33.00 ($25.00 for classroom use) ISBN: 0 913215 38 4

THE RIVERDALE COMPANY, Publishers
5506 Kenilworth Ave. - #102 - Riverdale, Maryland 20737 - (301) 864-2029

New from

Cambridge University Press

The Economics of Imperfect Information

Louis Phlips

Provides the first systematic textbook presentation of new microeconomic theories based on imperfect information. Each chapter explores a particular type of informational asymmetry and reviews the main contributions, from the seminal papers of the late 1960's through the most recent developments.

Paperback about $14.95

The Consequences of Economic Rhetoric

Edited by Arjo Klamer, Donald McCloskey and Robert M. Solow

This volume presents the results of a conference at Wellesley College, involving theoretical and applied economists, that explored the consequences of the rhetoric and the conversation of the field of economics.

Contributors include: Donald McCloskey, Arjo Klamer, Stanley Fish, Robert M. Solow, Robert L. Heilbroner, Rick Wolff, Stephen Resnick, A.W. Coats, Robert W. Clower, Cristina Bicchieri, Phil Mirowski, E. Roy Weintraub, Frank Denton, Nancy Folbre, Heidi Hartmann, Craufurd D. Goodwin, James Gailbraith, Robert O. Keohane, David Warsh.

Hardcover $39.50

Preference, Production and Capital

Selected Papers of Hirofumi Uzawa

Hirofumi Uzawa

Reviews various topics in economic theory, ranging from preference and consumption, duality and production, equilibrium, capital, and growth to the theory of social overhead capital.

Hardcover $49.50

Cournot Oligopoly

Characterization and Application

Edited by Andrew F. Daughety

Celebrating the 150th anniversary of Cournot's work, which Mark Blaug has characterized as "a book that for sheer originality and boldness of conception has no equal in the history of economics thought...", this volume focuses on the properties and uses of Cournot's model of competition among the few.

Contributors include: Andrew F. Daughety, Joseph Bertrand, John Nash, William Novshek, James W. Friedman, Roy Radner, Charles A. Holt, David M. Kreps, Jose A. Scheinkman, Xavier Vives, Paul Klemperer, Margaret Meyer, Hugo Sonnenschein, Esther Gal-Or, Thomas R. Palfrey, Stephen W. Salant, Sheldon Switzer, Robert J. Reynolds, John Vickers, Robert H. Porter, James A. Brander, Tracy R. Lewis.

Hardcover $54.50

Macroeconomics Policy Analysis

Open Economics with Quantity Constraints

Michael P. Amos

An introduction to the concept of public sector rationing in asset markets, this book provides a modern non-Walrasian framework that could serve as the basis for macroeconomic policy analysis.

Hardcover $27.95

New from
Cambridge University Press

Axiomatic Theory of Bargaining with a Variable Number of Agents

William Thomson and Terje Lensberg

An extrapolation on the Nash (1950) treatment of the bargaining problem to analyze the situation where the number of bargainers may vary, this book provides axioms to specify how solutions should respond to changes. It also provides new characterizations of all the major solutions as well as the generalizations of these solutions.

Hardcover $39.50

Statistical Games and Human Affairs

The View From Within

Roger J. Bowden

The book covers such problems as learning, publication, non-response, time inconsistency, and the existence of probabilities. Also included are discussions of attitude management, political polling and econometric modeling under heterogeneous information.

Hardcover $44.50

Macroeconomic Interactions Between North and South

Edited by David Currie and David Vines

These proceedings from a September 1987 conference organized by the Centre for Economic Policy Research and the International Economics Study Group focus on the strategic aspects of North-South interactions. The authors include some leading contributors to the growing literature on strategic interactions among interdependent economies.

Hardcover $49.50

Natural Monopoly Regulation

Principles and Practice

Sanford V. Berg and John Tschirhart

Develops a comprehensive framework for evaluation regulatory policies toward natural monopolies. After a brief historical overview of regulatory economics, the book provides analyses of optimal pricing and investment for natural monopolies and examines natural monopoly regulation in practice.

Cambridge Surveys of Economic Literature

About $59.50

International Economic Sanctions

Improving the Haphazard U.S. Legal Regime

Barry E. Carter

"This book presents a path-breaking analysis of the entire field of U.S. laws dealing with economic sanctions. It is distinguished in its comprehensive scope and as unique as a contribution to the scholarly literature." —***Phillip R. Trimble,*** UCLA Law School, Former Assistant Legal Adviser, Department of State

Hardcover $34.50

New from

Cambridge University Press

Optimal Control, Expectations and Uncertainty

Sean Holly and Andrew Hughes Hallet

The authors analyze the rational expectations revolution and discuss how the developments in economics, notably game theory, have greatly enhanced our understanding of how an economy functions and have made it possible for optimal control theory to be applied much more effectively to economic modeling and planning.

Hardcover about $54.50

Axioms of Cooperative Decision-Making

Herve Moulin

Discusses problems of fair division, equitable cost-sharing, division of a joint benefit or the choice of a truly democratic voting rule and other heterogeneous subjects linked by common axioms.

Hardcover $49.50

Socialist Planning

Second Edition

Michael Ellman

The text takes full account of decollectivization in China, perestroika and the Gorbachev reforms in the Soviet Union, the Polish economic crisis of 1979-82 and recent theoretical work by economists in Hungary, China and the Soviet Union.

Paperback about $19.95

Village Republics

Economic Conditions for Collective Action in South India

Robert Wade

Robert Wade's book is an important contribution to current debates about the problems of rural development in India and, more generally, about the capacity for collective action that exists within peasant societies.

Hardcover $44.50

OPEC

25 Years of Prices and Politics

Ian Skeet

This book looks at the history of OPEC, and the political and economic events that have shaped the organization and the world economy since its creation in 1960. It covers the background, the critical years of 1970-73 through its efforts to run a cartel in the 1980's.

Cambridge Energy Studies

Hardcover $29.95

Search Models and Applied Labor Economics

Nicholas M. Kiefer and George R. Neumann

Marks the development of empirical application of the search approach to labor economics. Included are econometric issues such as estimation and specification of search models for wages and unemployment duration, continuous time models of turnover, and identification of structural parameters.

Hardcover $47.50

New from
Cambridge University Press

Governing Economy
The Reformation of German Economic Discourse 1750-1840
Keith Tribe
This book reviews the distinctive features of the development of German economics and draws attention to its divergences from the Anglo-Saxon tradition.
Hardcover $44.50

Tax Reform in Disequilibrium Economics
Serge Wibaut
An analysis of the theory of optimal taxation within a disequilibrium framework, the book develops a model that is tested, then used to identify and numerically estimate desirable directions for tax reform in a specific macroeconomic context.
Hardcover about $39.50

A Microtheory of the Japanese Economy
Information, Incentives and Bargaining
Masahiko Aoki
Provides systematic descriptions of microeconomic institutions in the Japanese economy and interprets their workings in terms familiar to Western economists. The author critically examines some conventional notions about the microstructure of the market economy that have strongly shaped and influenced the economists' approach to industrial organization.
About $34.50

Conceptual Anomalies in Economics and Statistics
Lessons from the Social Experiment
Leland Gerson Neuberg
Explores anomalies in the conceptual basis of social control experiment and in the foundation of statistics and economics. Concludes that social experiments are an overrated tool for social policy improvements and that anomalies prevent microeconomics and statistics from providing a coherent understanding of human social behavior.
$54.50

Reduced Worktime and the Management of Production
Chris Nyland
Examines the tendency in market economies to reduce the time workers spend at their place of employment and considers the role scientific management has played in this development. Contends that the changing nature of worktime can be explained by changes in both the capitalistic production process and the demands that this process places on the psycho-physiological capacities of human beings.
About $29.95

At bookstores or order from
Cambridge University Press
32 East 57th Street, NY, NY 10022. Cambridge toll-free numbers for orders only:
800-872-7423, outside NY State, 800-227-0247, NY State only. MasterCard and Visa accepted.

ECONOMICS FROM CHICAGO

ALLOCATION OF INCOME WITHIN THE HOUSEHOLD

Edward P. Lazear *and* Robert T. Mitchell

Lazear and Michael develop an important new economic model of income distribution within the family, one that attempts to determine which family characteristics affect spending patterns. They also explore their study's implications for the process of determining child support payments in households that dissolve.

Cloth $34.95 228 pages tables throughout

MISALIGNMENT OF EXCHANGE RATES

Effects on Trade and Industry

***Edited by* Richard C. Marston**

Contributors investigate the causes of misalignment and the extent to which observed movements in real exchange rates can be attributed to misalignment. These studies are conducted both empirically, through the experiences of several countries, and theoretically, through models of imperfect competition.

Cloth $37.50 336 pages 66 line drawings, 52 tables
An NBER Project Report

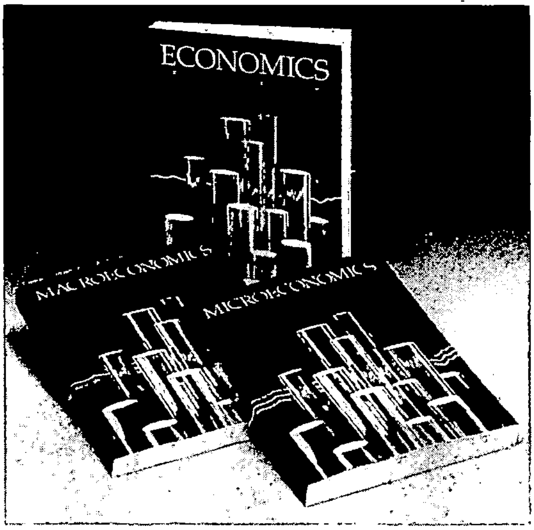
CHICAGO STUDIES IN POLITICAL ECONOMY

***Edited by* George J. Stigler**

"There is no question that a well-defined 'Chicago School' of political economy has emerged, built largely around the work of George J. Stigler and his colleagues. *Chicago Studies in Political Economy* brings together the key works in this field, works that have been extremely influential among economists who study political processes. It is a collection of enormous value."
--Roger G. Noll, Stanford University

Paper $19.95 664 pages tables throughout
Library cloth edition $60.00

THE UNIVERSITY OF CHICAGO PRESS
5801 South Ellis Avenue, Chicago, IL 60637



THE RATIONAL CHOICE IS THE NATIONAL CHOICE

Economics
Second Edition
Martin Bronfenbrenner
Aoyama Gakuin University, Japan
Emeritus, Duke University
Werner Sichel and **Wayland Gardner**
Both of Western Michigan University
Complete Hardcover Edition • 828 pages
Two-volume Paperback Edition:
Macroeconomics • 460 pages
Microeconomics • 556 pages
Study Guide and Test Bank by **Rose Pfefferbaum,** Mesa Community College
Instructor's Manual • Transparency Masters
Color Transparencies • MicroStudy: Computerized Study Guide • MicroTest: Computerized Test Bank • Call-in Test Service • Color Graphics Software
GPA: Grade Performance Analyzer • PC Economics: Interactive Graphing Software and Workbook • 1987

For adoption consideration, request an examination package from your regional Houghton Mifflin office.

Houghton Mifflin

13400 Midway Rd., Dallas, TX 75244-5165
1900 S. Batavia Ave., Geneva, IL 60134
925 E. Meadow Dr., Palo Alto, CA 94303
101 Campus Dr., Princeton, NJ 08540

Timely Economic Analyses from Lexington Books

The Enlightened Society
The Economy with a Human Face
Kurt F. Flexner, Memphis State University
Foreword by Lawrence R. Klein

In this thought-provoking and readable book, Flexner searches for a more realistic understanding of the economic process within the total framework of human activity.
0-669-19930-3 336 pages
December 1988 tent. $39.95

Issues in Pricing
Theory and Research
Timothy M. Devinney, Vanderbilt University, editor

This comprehensive volume examines important new research and theory in the applied field of pricing strategy, in which pricing is viewed as a strategic decision variable.
0-669-13215-2 432 pages 1988 $55.00

Food Processing An Industrial Powerhouse in Transition
John M. Connor, Purdue University

This outstanding new book thoroughly examines the growth and economic development of U.S. commercial food processing activity from 1963-85.
0-669-19511-1 496 pages 1988 $50.00

Recent Developments in International Banking and Finance Volume II
Sarkis J. Khoury, University of California, Riverside, and Alo Ghosh, McKinsey & Co., editors

A dynamic look at the latest issues, ideas and problems in international banking and finance, providing thoroughly up-to-date coverage of critical new and continuing areas of concern.
0-669-16197-7 432 pages 1988 $59.00

Lexington Books
125 Spring Street
Lexington, MA 02173
Call Toll Free
(800) 235-3565

DC Heath
A Raytheon Company

Beyond Free Markets
The Revival of Activist Economics
Marc Levinson

In a clear, non-technical presentation, Levinson explains the shortcomings of previous policies and makes a rigorous case for applying new activist ideas to the most pressing problems facing the U.S. and world economics.
0-669-16971-4 224 pages 1988 $19.95

Cooperative Research in U.S. Manufacturing
Assessing Policy Initiatives and Corporate Strategies
Albert N. Link, University of North Carolina at Greensboro, and Laura L. Bauer, Quick, Finan and Associates

This ground-breaking book is the first to analyze cooperative research among firms: what types of firms participate in these research relationships; their motives; and the economic consequences of cooperative endeavors.
0-669-16969-2 160 pages
December 1988 tent. $29.00

Japan's Economic Strategy in Brazil
Challenge for the United States
Leon Hollerman, St. Olaif College

In this innovative and perceptive book, noted Japanologist Leon Hollerman uses Japan's partnership with Brazil as a case study of Japan's challenge to the leadership and competitive position of the U.S. within a third country.
0-669-17428-9 304 pages 1988 $41.00

Transnational Corporations and China's Open Door Policy
Teng Weizao, Nankai University, China, and N.T. Wang, Columbia University, editors

This comprehensive volume explores important issues, theories and experiences which illuminate the emerging symbiosis between transnational corporations and China, in the new era of its Open Policy.
0-669-16967-6 304 pages 1988 $45.00

For IBM Personal Computers

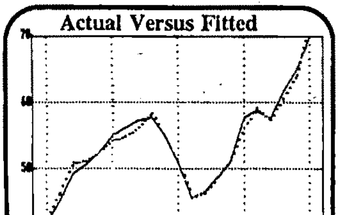
ySTAT™

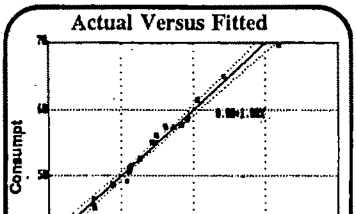
Integrates
SPREADSHEET + STATISTICS + FORECASTING + GRAPHICS
and is Easy to Use.

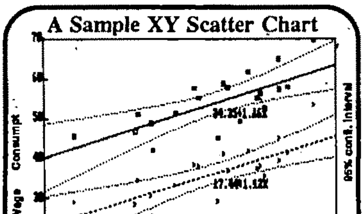
- ySTAT is the only full-fledged spreadsheet statistical program - a fast, powerful, easy-to-use stand-alone program.
- **Menu-driven** format as in Lotus 1-2-3 - offers many functions that are not available even in 1-2-3, such as vector formulas, lagged variables, dummy variables, moving averages, and interpolation of missing values.
- You can read, enter, edit, copy, move, and sort the data as well as group and partition the data for analysis right on the spreadsheet. A total of 40 function keys for one-keystroke operations.
- **Ten statistical function keys:** to tally sums and means; subtract variable means or seasonal means; standardize the data; aggregate seasonal data; generate seasonal indicators; perform moving averages, linear trend forecasts, and so on.
- **ySTAT reads 1-2-3 and Symphony worksheet files directly,** and any free-format or fixed-format textfile generated by other PC programs or downloaded from mainframe. The data can be numeric, alphabetic, including missing values.
- Mean, median, quartiles, skewness, mode, t-test, frequency distribution, correlation, crosstabs, analysis of variance.
- Multiple **regression** (OLS), standardized regression, weighted least squares, two-stage least squares, polynomial regression, and Cochrane-Orcutt method. Including diagnostic residual analysis output (see sample at right).
- **Pooling** of cross-section and time-series data.
- Two-sample difference of means test, **nonparametric tests** (Kolmogorov-Smirnov test, Wald-Wolfowitz runs test, Mann-Whitney test, Wilcoxon matched-pairs signed-ranks test) and Spearmans's and Kendall's rank-order correlations.
- **Time-series models:** autoregressive model, moving average model, and Box-Jenkins model.
- **Nonseasonal forecasts:** linear trend, quadratic trend, polynomial model, sinusoidal model; simple, double and triple exponential smoothings.
- **Seasonal forecasts:** exponential smoothings with seasonal indicators or with trigonometric functions; Winters' additive and multiplicative procedures.
- **LOGIT, PROBIT, TOBIT,** and Weighted Probit regressions: these models are twice as fast as Gauss program and faster than other programs and are much easier to use.

High-resolution **color charts** : the actual and forecast values are graphically displayed as a line chart and also as an XY chart, including a regression line and its 95% confidence interval.

```
                    OLS  -- DEPENDENT VARIABLE: Consumpt

   RIGHT-HAND          ESTIMATED         STANDARD           T_STATISTIC       PROB.
    VARIABLE          COEFFICIENT          ERROR

  1  Profits          0.192934381    (      0.09121)    T=      2.11527      0.049
  2       P-1         0.089864898    (      0.09065)    T=      0.99158      0.335
  3      W+W'         0.796218750    (      0.03994)    T=     19.93342      0.000
  4 Constant         16.236600272    (      1.30270)    T=     12.46382      0.000

 SAMPLE SIZE(  1 to   21) =             21        (DF=17)
 SUM OF SQUARED RESIDUALS =             17.879449
           VARIANCE (MSE) =              1.051732
STANDARD ERROR (ROOT MSE) =              1.025540
                R-SQUARED =              0.981008
       ADJUSTED R-SQUARED =              0.977657
  F-STATISTIC(  3,    17) =            292.707595 (p=0.0004)
         SUM OF RESIDUALS =             -0.000000
  DURBIN-WATSON STATISTIC =              1.367474

Analysis of Variance|              Source |      SUM SQ     | DF |   MEAN SQ   |
                    |---------------------|-----------------|----|-------------|
                    |Due to Regression    |         923.550 |  3 |     307.850 |
                    |            Residual |          17.879 | 17 |       1.052 |
                    |               Total |         941.430 | 20 |      47.071 |
                    |---------------------|-----------------|----|-------------|

RESIDUAL ANALYSIS -Mean: -0.000   Adj. RMSE: 0.946   Mean Abs. % Err: 1.283

                              AUTOCORRELATIONS

LAG    COEF     T-VAL   -+----+----+----+----+----0----+----+----+----+----+-  LAG
  1  0.181      0.83   |                          |>>>>>>>>>>>               |    1
  2 -0.059     -0.27   |                       <<<|                          |    2
  3 -0.033     -0.15   |                        <<|                          |    3
  4 -0.157     -0.69   |                 <<<<<<<<<|                          |    4
  5 -0.411     -1.81   |<<<<<<<<<<<<<<<<<<<<<<<<<<|                          |    5
                        -+----+----+----+----+----0----+----+----+----+----+-  LAG
    Ljung-Box statistic (chi-square 4 DF): 6.698 (p=0.1527)

                           PARTIAL AUTOCORRELATION

LAG    COEF     T-VAL   -+----+----+----+----+----0----+----+----+----+----+-  LAG
  1  0.181      0.83   |                          |>>>>>>>>>>>               |    1
  2 -0.095     -0.44   |                    <<<<<<|                          |    2
  3 -0.004     -0.02   |                          |                          |    3
  4 -0.162     -0.74   |                <<<<<<<<<<|                          |    4
  5 -0.378     -1.73   |<<<<<<<<<<<<<<<<<<<<<<<<<<|                          |    5
                        -+----+----+----+----+----0----+----+----+----+----+-  LAG

                   ACTUAL versus FITTED VALUES and RESIDUALS

SEQ      Actual    Fitted min=41.900                   71.873=max  Error
          (*)       (+)   -+----+----+----+----+----+----+----+-              +---+---+
  1      41.90     42.22 |#                                        |  -0.324  |  <:    |
  2      45.00     46.25 |    *+                                   |  -1.250  |<<<:    |
  3      49.20     50.77 |          * +                            |  -1.566  |<<<:    |
  4      50.60     51.09 |            #                            |  -0.494  |  <:    |
  5      52.60     52.59 |              #                          |   0.008  |   :>   |
  6      55.10     54.23 |                +*                       |   0.869  |   :>>  |
  7      56.20     54.86 |                +  *                     |   1.338  |   :>>> |
  8      57.30     56.25 |                   +*                    |   1.055  |   :>>  |
  9      57.80     58.39 |                    *+                   |  -0.589  | <<:    |
 10      55.00     54.72 |                 #                       |   0.282  |   :>   |
 11      50.90     51.13 |            #                            |  -0.230  |  <:    |
 12      45.60     45.92 |     *+                                  |  -0.322  |  <:    |
 13      46.50     46.18 |      #                                  |   0.322  |   :>   |
 14      48.70     48.76 |         #                               |  -0.058  |  <:    |
 15      51.30     51.33 |             #                           |  -0.035  |  <:    |
 16      57.70     56.08 |                   +*                    |   1.616  |   :>>> |
 17      58.70     59.14 |                      *+                 |  -0.436  |  <:    |
 18      57.50     57.29 |                     #                   |   0.210  |   :>   |
 19      61.60     60.61 |                         +  *            |   0.989  |   :>>  |
 20      65.00     64.21 |                              +*         |   0.785  |   :>>  |
 21      69.70     71.87 |                                     *  +|  -2.173  <<<<:    |
          (*)       (+)   -+----+----+----+----+----+----+----+-              +---+---+

            LIST OF OUTLIERS - STUDENTIZED RESIDUALS GREATER THAN 1.50

   SEQ      Actual    Fitted     Residual   Std. Err  Studentized Probability
                                                        Residuals     DF=17
   ---      ------    ------     --------   --------  -----------  ---------
     3       49.20     50.77       -1.566      0.967       -1.619      0.124
    16       57.70     56.08        1.616      0.958        1.687      0.110
    21       69.70     71.87       -2.173      0.780       -2.787      0.013
   ---      ------    ------     --------   --------  -----------  ---------

[Note: this sample output shows some of the options for regression diagnostics.]
```


Actual Versus Fitted


Actual Versus Fitted


A Sample XY Scatter Chart

MING TELECOMPUTING INC.

23 Oak Meadow Road, P.O. Box 101, Lincoln Center, MA 01773, U.S.A. (617) 259-0391

Order: ( ) ySTAT for $395. ( )ySTAT/Medical for $395.
( ) Trial Disk and information/sample output for $5.
( ) Information/sample output.
System: ( ) IBM PC. ( )XT. ( )AT. ( )System 2 Model____
( ) Other system ______________________
Options: ( ) with 8087 or 80287 or 80387 math co-processor.
( ) IBM CGA ( ) EGA ( ) Hercules Mono. Card
Type of printer or plotter ______________________

Payment: ( ) Check or money order.
( ) U.S. university or governmental purchase order.
( ) Visa. ( ) MC. Card No. ______________________
Name ______________________ Expir. date ___/___
Address ______________________
______________________ Telephone: ( ) ________

IBM is a registered trademark of International business Machines Corp.; 1-2-3 of Lotus Development Corp.; Gauss of Aptech Systems, Inc.

DRYDENOMICS

The Best of 1988 and 1989

ECONOMICS: THEORY AND PRACTICE, 3/E
Patrick J. Welch and Gerry F. Welch

ECONOMICS, 5/E
(also available in Macro and Micro volumes)
Edwin G. Dolan and David E. Lindsey

MICROECONOMIC THEORY: BASIC PRINCIPLES AND EXTENSIONS
Walter Nicholson

FUNDAMENTALS OF MANAGERIAL ECONOMICS, 3/E
James Pappas and Mark Hirschey

THE PRICE SYSTEM AND RESOURCE ALLOCATION, 10/E
Ross D. Eckert and Richard H. Leftwich

THE ECONOMICS OF LABOR MARKETS AND LABOR RELATIONS, 2/E
Bruce E. Kaufman

THE WORLD ECONOMY: TRADE AND FINANCE
Beth V. Yarbrough and Robert M. Yarbrough

GOVERNMENT AND THE MARKETPLACE, 2/E
Peter Asch and Rosalind S. Seneca

THE ANTITRUST CASEBOOK: MILESTONES IN ECONOMIC REGULATION, 2/E
William Breit and Kenneth Elzinga

COMPARATIVE ECONOMIC SYSTEMS
H. Stephen Gardner

AMERICAN ECONOMIC HISTORY, 2/E
Robert C. Puth

MACROECONOMIC DECISION MAKING IN THE WORLD ECONOMY: TEXT AND CASES, 2/E
Michael G. Rukstad

MONEY AND BANKING: A MARKET-ORIENTED APPROACH, 3/E
Ivan C. Johnson and William W. Roberts

MONEY, BANKING AND MONETARY POLICY
Colin D. Campbell, Rosemary G. Campbell, and Edwin G. Dolan

THE DRYDEN PRESS
908 North Elm Street, Suite 101, Hinsdale, IL 60521

D61-8AER-F

Announcing a New Journal and Call for Papers

Editor
SOON CHO
Seoul National University

Co-editor
SHIN-HAING KIM
Seoul National University

Board of Editors
HYUNG-YOON BYUN
BYUNG-KWON CHA
SEUNG-SOO HAN
WONTACK HONG
KI-JUN JEONG
JAE-YOON PARK
Seoul National University
GARY S. BECKER
University of Chicago
JEAN C. BENARD
University of Paris I
JAMES A. MIRRLEES
Oxford University
TAKASHI NEGISHI
University of Tokyo
JEFFREY D. SACHS
Harvard University

in association with Institute of Economic Research, Seoul National University

Director
WOO-HEE PARK

SEOUL JOURNAL OF ECONOMICS

Seoul Journal of Economics is designed to provide an outlet for research on all aspects of economic development (covering cultural, historical, and sociological approaches), specially focusing on the economic development of East Asia. The journal also seeks to publish recent theoretical developments in all specialized fields of economics as well as rigorous empirical works.

Some of the papers featured in the first volume are

Ronald I. McKinnon and Kenichi Ohno
Getting the Exchange Rate Right: Insular vs. Open Economies

Myoung-Kyu Kang
The Growth of Economics Profession in Korea.

Paul Burrows and Gary Yohe
Monopoly Power and the Optimal Control of External Costs

Jeffrey D. Sachs
Prospects for Global Trade Imbalances: A Simulation Approach

Submission of Papers should be addressed to Editor, Seoul Journal of Economics, Institute of Economic Research, Seoul National University, Seoul 151-742, Korea

Subscription rate:
☐ Individuals US $20.00 ☐ Institutions US $40.00

ORDER FORM

Payment is enclosed-Cheque ( ) Money Order ( )

Address: Street no.

City State

To order please mail this coupon.
Make cheque or money order payable to

Seoul Journal of Economics
INSTITUTE OF ECONOMIC RESEARCH
SEOUL NATIONAL UNIVERSITY
Seoul 151-742, Korea

GENERAL EQUILIBRIUM MODELS
FOR MICRO-COMPUTERS

GEMODEL.USA

A large-scale general equilibrium model of the U.S. economy

The package is completely menu-driven and accomplishes every job from data entry through consistency checks, calibration and policy changes to solution and report printing.

The model can be used to analyze the effects of changes in tax policy, product demands, technology, productivity etc. on incomes, outputs, tax revenue and foreign trade.

GEMODEL.USA
is for policy analysis
and research in
government, university
and industry.

A special purpose of
the model is
to simulate
value-added taxes
and the separate tax
policies of two
levels of government:
Federal
State/Local

GEMODEL.USA has a phenomenal number-crunching capability. It accepts the 85-industry, U.S. input-output tables prior to aggregation of industries, consumption categories, and households. The level of detail carried by the model is larger than that commonly found in the general equilibrium literature.

The package is easy to run and is perhaps the only software that produces simulation results on your own desk within seconds. It is available in two versions: Professional and Academic.

PRICE:
Academic Version US$ 2,200

SYSTEM REQUIREMENTS: IBM-PC or compatible, 256K RAM, 2 drives, DOS 2.1 or better.

GEMODEL 2.0

Solves open or closed, two or three-industry small open economy models employing two or three factors. Industries can have constant or diminishing returns to scale. Admits one to nineteen household groups with or without income/leisure choice. Solves with a wide array of factor, commodity and income taxes.

PRICE: US$ 395.00

Ideal to enrich
courses in Price
Theory, Trade and
Finance by solving
models
within seconds
at 48 different
levels of complexity
with
18 to 176 equations.

GESTATS

Calibrates GEMODEL 2.0 parameters to data and elasticity assumptions. Checks data consistency.
PRICE: US$ 195.00

GEREPORT

Compares GEMODEL 2.0 simulation results with the base case or with other simulation results. Computes measures of welfare change and marginal costs of taxes.
PRICE: US$ 95.00

GEDATA

18 data files for exercises in price theory, trade and finance using GEMODEL 2.0
PRICE: US$ 25.00

For information and to order, write to:
DIA Agency Inc., 1879 Kingsdale Ave., Ottawa, Ontario, K1T 1H9 CANADA
DEMONSTRATOR diskettes available at $25.

Solid Principles Texts The 1989 Economic Outlook

from South-Western

PRINCIPLES OF ECONOMICS, 4th Edition

Ryan C. Amacher and Holley H. Ulbrich
Clemson University

This text's pragmatic approach and balanced presentation of principles teaches students the "economic way of thinking" in everyday life. Significant revisions to this edition include: incorporation of international perspectives throughout the text, and new appendices on farm and urban problems and on indifference analysis. The new package also includes a completely new Study Guide, an Instructor's Manual, and a video covering several topics in economics. The text is available in comprehensive format or in separate volumes: **Principles of Microeconomics** and **Principles of Macroeconomics**.

ECONOMICS: The Science of Common Sense, 6th Edition

Elbert V. Bowden, Chair of Banking,
Appalachian State University

The straightforward language and logical organization of this text help students grasp concepts more quickly. New additions to this popular text include: two new chapters on the evolution of macrotheory; a chapter on the Keynesian/Monetarist debate, presented from a fresh perspective; updated information on the conditions, events, and problems of the late 1980s, including the stock market crash and deficits. The text is available in both comprehensive and abridged editions.

For more information about the texts that shape South-Western's 1989 Economic Outlook, contact your area representative or write:

SOUTH-WESTERN
COLLEGE DIVISION

5101 Madison Road
Cincinnati, OH 45227

American Economic Association/Federal Reserve System Minority Graduate Fellowships in Economics

The American Economic Association and the Federal Reserve System are pleased to announce their joint sponsorship of graduate fellowships to minority Ph.D. students who have completed their comprehensive examinations and, if applicable, their field examinations, and are about to begin their dissertation research. Awards will be based on academic performance.

Applicants must be U.S. citizens who are Black, Hispanic or Native American and are enrolled in an accredited graduate program in Economics in the United States. Preference will be given to applicants whose area of concentration is of special interest to the Federal Reserve System (e.g., financial markets and monetary policy, nonfinancial macroeconomics, forecasting, banking markets and financial structure, regional studies, the external sector of the U.S. economy, the economies of other countries, foreign exchange markets, and international banking and financial markets).

A stipend of $700 per month for the academic year and tuition relief by the institution nominating the student and verifying successful participation in its graduate program in Economics are provided. Recipients will be assigned an adviser from the Federal Reserve System and be given the opportunity to work for one summer at the Board or a Federal Reserve Bank.

Applications are due March 1, 1989
Awards will be announced early in April 1989

For further information and application materials write:

Barbara Sears, Registrar
AEA/FRS Minority Fellowship Program
Department of Economics
College of Business and Public Administration
University of Arizona
Tucson, AZ 85721
(602) 887-5887

KLASSIKER DER NATIONALÖKONOMIE

Die Handelsblatt-Bibliothek.

»A scientific and literary event of secular rank«

Horst Claus Recktenwald

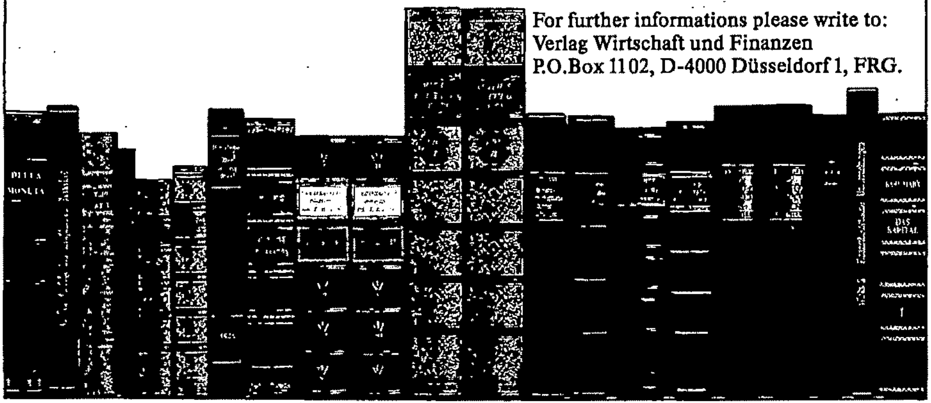
- The »Klassiker der Nationalökonomie« – facsimile-edition opens access to a treasure of economic science and culture.
- Stimulating praefatio (with archival witnesses) and facsimiles of first editions are a true and sure vademecum for daily users as for bibliophiles.
- Outstanding experts answer the fascinating question what is still classic in our epoch by critically analysing rare masterpieces down the ages.
- »Klassiker der Nationalökonomie« is an unique opportunity to build up a valuable private library sui generis.

»The »Klassiker der Nationalökonomie« undertaking to publish reproductions of a wide selection of economic classics in their first editions is unprecedented. It is equally remarkable that the reproductions are presented in handsome, durable form, and accompanied by authoritative essays by contemporary economists.«

George J. Stigler

»With renewed interest in the history of economic thought, the publication of »Klassiker der Nationalökonomie« seems to be particularly timely. The luxurious facsimile edition will delight not only professional economists but also bibliophiles attracted by the exemplary printing and binding.«

Wassily Leontief



For further informations please write to:
Verlag Wirtschaft und Finanzen
P.O.Box 1102, D-4000 Düsseldorf 1, FRG.

INDEX OF ECONOMIC ARTICLES

prepared by

The Journal of Economic Literature
of the
American Economic Association

✓ Each volume in the **Index** lists articles in major economic journals and in collective volumes published during a specific year.

✓ No other single reference source covers as many articles classified in economic categories as the **Index.**

Index volumes XI-XXII covering 1969-1980 are available at $60.00 each.

The following two part **Index** volumes are now ready for delivery at $90.00 per set:

Volume	Year
XXIII	1981
XXIV	1982
XXV	1983
XXVI	1984

Note: 30% discount for AEA members (individuals only); 15% discount for libraries and subscribers.

an indispensable tool for...
ECONOMISTS
REFERENCE LIBRARIANS
RESEARCHERS
TEACHERS
STUDENTS
AUTHORS

Payment required in advance. Prices include shipping charges; allow 4-6 weeks for delivery. Please send your check or money order (net of applicable discount) payable in United States dollars drawn on a United States bank to:

American Economic Assn. - **Index**
P.O. Box 20111
Nashville, TN 37202

Address inquiries or other correspondence to:
Journal of Economic Literature, P.O. Box 7320
Oakland Station, Pittsburgh, PA 15213.

Begin making plans to attend the

Annual Meeting of The American Economic Association

(in Conjunction with Allied Social Science Associations)

to be held in

NEW YORK, NY

Dec. 27-30, 1988

The Employment Center opens Tuesday, December 27; sessions begin Wednesday, December 28.

See the September *AER* for the American Economic Association's preliminary program.

The 1989 meeting will be held in Atlanta, GA, December 27-30.

JOB OPENINGS FOR ECONOMISTS

Available only to AEA members and institutions that agree to list their openings.

Annual Subscription Rates

U.S.A., Canada, and Mexico (first class):	$15.00, regular AEA members and institutions $ 7.50, junior members of AEA
All other countries (air mail):	$22.50, regular AEA members and institutions $15.00, junior members of AEA

Please begin my issues with:

☐ February ☐ April ☐ June ☐ August ☐ October ☐ December

Name____
First Middle Last

Address______________________________________

____
City State/Country Zip/Postal Code

Check one:

☐ I am a member of the American Economic Association.
☐ I would like to become a member. My application and payment are enclosed.
☐ (For institutions) We agree to list our vacancies in JOE.

Send payment (U.S. currency only) to:

THE AMERICAN ECONOMIC ASSOCIATION
1313 21st Avenue South
Nashville, Tennessee 37212

ECAAR

Economists Against the Arms Race

A new organization dedicated to preventing war
and fostering global cooperation

Board of Trustees

Kenneth J. Arrow, *co-chair* Lawrence R. Klein, *co-chair*
Wassily Leontief, Franco Modigliani, Robert M. Solow, James Tobin

"Warfare affects human welfare to such an extent that economists must raise their voices against it." —Jan Tinbergen

Planned Activities:

- Organization of discussions and conferences
- Lectures and public debates
- Formation of member chapters at colleges and universities
- Writing, publication and dissemination of research results
- Books and articles to inform decision-makers and the public

GOING PUBLIC
Economists Against the Arms Race
Friday, December 30 at 8:00 AM Continental Breakfast
the Green Room New York Hilton Hotel

You are cordially invited by the Trustees
and **Robert Eisner, Joseph Peckman, John Kenneth Galbraith, Dietrich Fischer** and **Robert Schwartz**
to join them in launching **ECAAR.**

See our table at the AEA Convention!

For more information and membership application, please write to
ECAAR, 30 West 95th St, New York, NY 10025,
attn: Margaret Hartnett.

Journal of International Economic Integration

Solicits Papers to Compete for the Annual Daeyang Prize in Economics ($7,000) and Welcomes Subscriptions by Interested Parties

Current Issues include

Bela Balassa, *Japanese Trade Policies Towards Developing Countries.*

Basant K. Kapur, *Open-Economy Response to a Terms of Trade Shock in a Growth Context.*

Norman C. Miller, *A Gerneral Approach to the Balance of Payments and Exchange Rates.*

Leonard F.S. Wang, *Product Market Imperfections and Customs Unions theory.*

Gene M. Grossman, *The Employment and Wage Effects of Import Competition in the United States.*

Wilfred J. Ethler and Ronald D. Fischer, *The New Protectionism*

Joshua Alzenman, *Inflation, Tariffs and Tax Enforcement Costs*

Chung H. Lee and Seljl Naya, *The Internationalization of U.S. Service Industries and Its Implications for Developing Countries.*

The Journal of International Economic Integration is published biannually (Spring and Autumn) by the Institute for International Economics, King Sejong University, Seoul, Korea.
The purpose of the Journal is to support and encourage research in the area of international trade, international finance and other related economic issues that include general professional interest in international economic affairs. Welcoming both theoretical and empirical analyses in international economics, the Journal is strongly interested in the issues of the international economic cooperation.

- The Journal welcomes unsolicited manuscripts, which will be considered for publication by the Editorial Board.
- From papers selected for publication, the Prize committee will choose the best manuscript(s) to receive the $7,000 Daeyang Prize. The winner of the prize is announced in the Spring issue every year.
- The manuscripts should be accompanied by an abstract of no more than 100 words and a brief curriculum vitae containing the author's academic career. All submissions should be typewritten, double-spaced, in English with footnotes, references, figures, tables and any other illustrative material on separate sheets.
- Three copies of the manuscript and all accompanying material should be submitted to the following address by October 31, 1988 for consideration for 1989 publication.
- For subscriptions to the Journal ($20 per year for individuals, $30 per year for institutions), send a check or money order payable to King Sejong University to the following address.

Institute for International Economics
King Sejong University
Seongdong-Ku, Seoul, Korea

OECD OCDE

Industrial Revival Through Technology.

New technologies are permeating industry at an increasing speed, altering production techniques and giving rise to new products. This report examines the extent to which these technologies can contribute to revitalizing manufacturing activities that have exhibited only moderate expansion or contraction. Seven industries have been chosen to illustrate the role of new technologies for industrial revival: printing and publishing, glass, textile, clothing, food processing, watch, and paper industries.

70-88-02-1, September 1988, 228 pages, ISBN 92-64-13103-5, $22.00

The Newly Industrialising Countries: Challenge and Opportunity for OECD Industries.

The Newly Industrializing Countries have become important producers and exporters of manufactured goods, moving up from traditional to increasingly sophisticated products. By 1983, the OECD area's long standing trade surplus in manufactures with the six NICs considered here—South Korea, Taiwan, Singapore, Hong Kong, Mexico, and Brazil—had turned into a deficit which rapidly attained nearly 18 billion dollars. This report documents the evolution of this trade over the past two decades, focusing on the industrial policy and adjustment issues which it raises. It examines the "double dynamic" of indigenous impetus to industrialization associated with foreign investment and technology transfer. It also surveys the challenges and opportunities for OECD economies as they seek to strengthen their own competitiveness while intensifying industrial and technological cooperation with the NICs.

70-88-01-1, July 1988, 151 pages, ISBN 92-64-13041-1, $20.00

Tax Consequences of Foreign Exchange Gains and Losses.

Issues in International Taxation Series No. 3. Discusses the implications of fluctuations in foreign exchange rates on the taxation of the ordinary income and capital gains of enterprises and reviews country practices.

23-88-06-1, September 1988, 71 pages, ISBN 92-64-13124-8, $22.00

Agricultural Policies, Markets and Trade:
Monitoring and Outlook 1988.

The first annual monitoring and outlook report on agricultural policies, markets, and trade in OECD Member countries. It analyzes recent developments (up until early 1988) in the light of the principles for agricultural policy reform outlined by the 1987 Ministerial Council. It also provides data on the levels at which governments have subsidized agriculture for the years 1979-1986, data on production, consumption, and trade of various commodities, and information on the impact of developing countries, Eastern Europe, the USSR, and China on the world agricultural market.

51-88-03-1, September 1988, 228 pages, ISBN 92-64-13126-4, $22.00

To order, send your check or money order to:

OECD Publications and Information Center
2001 L Street, NW, Washington, DC 20036-4095
Telephone: (202) 785-6323

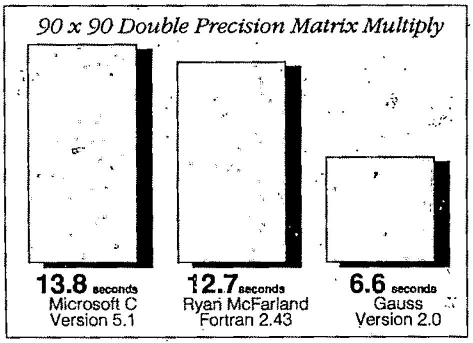
GAUSS RUN WITH THE BEST

Saves time. Saves money. New GAUSS 2.0.

Now you can run most routines at or near the capacity of your hardware.

From Bell Labs to CalTech, research scientists and engineers have found GAUSS can help them solve problems with their PCs that would take them hours longer wth conventional languages. And not only does GAUSS run fast, it's easier to program because its language is closer to mathematical syntax. GAUSS allows you to easily enter your own ideas even if you are not a programmer and saves time by including over 250 mathematical routines. You can even call your favorite Fortran, C or assembler subroutines.

90 x 90 Double Precision Matrix Multiply



Now GAUSS 2.0, the second generation of the new breed of matrix handling programming languages is available, with indexed, easier-to-use documentation, on-line debugger and Smart™ on-line help that can even handle your own functions. And GAUSS 2.0 is faster, up to 20% on some calculations, compared to our previous versions.

The GAUSS system includes publication quality graphics and a library of mathematical routines and applications, documented in the updated 2.0 manual. All this for the total price of just $395. So don't waste another minute running at half speed, order GAUSS 2.0 today.

We're so sure GAUSS will make you more productive, giving you more answers to more questions more quickly, that we guarantee it.

If you are not completely satisfied with the power, speed and flexibility of GAUSS simply return it for a full refund. No questions asked. Call **(206) 631-6679** to order.

GAUSS - faster, easier, smarter...guaranteed.

SATISFACTION • GUARANTEE • CUSTOMER
100%

GAUSS 2.0

FOR IBM PC-XT-AT-PS/2 & COMPATIBLES
WITH MATH COPROCESSOR
BY APTECH SYSTEMS, INC.
26250 - 196th Place SE
Kent Washington 98042
Phone (206) 631-6679
FAX (206) 630-1220
CALL OR WRITE FOR FREE BROCHURE

IBM IS A REGISTERED TRADEMARK OF INTERNATIONAL BUSINESS MACHINES
ALL TESTS RUN ON AN IBM PS/2 MODEL 80 16 MHz